

Université de Blida 1 – SAAD DAHLAB



Faculté des sciences

Département d'Informatique

Mémoire présenté par :

Mlles.

DJIDA Chahrazed

et YAMI Safia

Pour l'obtention du diplôme de Master

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Traitement Automatique de la Langue

Sujet :

**Sentiment Analysis of students reviews
on ELearning platforms**

Defended on: 06/07/2020, in front of the members of the thesis committee:

Mrs. S. OUKID	Professor	University of Blida 1	President
Mr. S. FERFERA	Associate Professor	University of Blida 1	Examinator
Mrs. M. MEZZI	Associate professor	University of Blida 1	Supervisor

Abstract

E-learning is becoming one of the most effective training approaches nowadays. Thanks to the E-Learning platforms and their collaboration tools, students can interact with each other and share their doubts on certain subjects through online reviews and forums.

Nowadays E-Learning platforms endow for students to review courses. However, teachers often remain outside this process and are sometimes, not aware about the learning problems encountered in their classes. The solution could be adopting a sentiment analysis methodology to students reviews in order to detect the mood and also the issues during the learning process. That will guarantee problem solving and the confidentiality of communications between students. Thus, affecting beneficially the students experience as a whole in E-Learning platforms

In this thesis we demonstrated the attempt of using sentiment analysis on E-Learning reviews to detect E-Learning students problems. In particular we focused on proposing a feedback method guaranteeing the confidentiality of communication and helping to improve of the E-Learning experience.

Keywords: *E-Learning, Sentiment Analysis, Opinion mining, Machine Learning, Data Processing, Data Filtration.*

Résumé

L'apprentissage en ligne est en train de devenir l'une des approches de formation les plus efficaces de nos jours. Grâce aux plateformes E-Learning et à leurs outils de collaboration, les étudiants peuvent interagir les uns avec les autres et partager leurs doutes sur certains sujets à travers des avis et des forums en ligne.

De nos jours, les plates-formes d'apprentissage en ligne permettent aux étudiants de revoir les cours. Cependant, les enseignants restent souvent en dehors de ce processus et ne sont parfois pas conscients des problèmes d'apprentissage rencontrés dans leurs classes. La solution pourrait être l'adoption d'une méthodologie d'analyse des sentiments pour les critiques des étudiants afin de détecter l'humeur et les problèmes pendant le processus d'apprentissage. Cela garantira la résolution de problèmes et la confidentialité des communications entre les étudiants. Ainsi, affectant de manière bénéfique l'expérience des étudiants dans son ensemble dans les plateformes de formation en ligne

Dans cette thèse, nous avons démontré la tentative d'utiliser l'analyse des sentiments sur les évaluations E-Learning pour détecter les problèmes des étudiants E-Learning. Nous nous sommes notamment attachés à proposer une méthode de retour d'expérience garantissant la confidentialité de la communication et contribuant à l'amélioration de l'expérience E-Learning.

Mots clés : *Apprentissage en ligne, Analyse des Sentiments, Analyse d'Opinion, Apprentissage Automatique, Traitement des Données, Filtration des Données.*

ملخص

أصبح التعلم الإلكتروني أحد أكثر مناهج التدريب فعالية في الوقت الحاضر. بفضل منصات التعلم الإلكتروني وأدوات التعاون الخاصة بهم، يمكن للطلاب التفاعل مع بعضهم البعض ومشاركة شكوهم حول مواضيع معينة من خلال المراجعات والمننديات عبر الإنترنت.

توفر منصات التعلم الإلكتروني في الوقت الحاضر للطلاب إمكانية مراجعة الدورات. ومع ذلك ، غالبًا ما يظل المعلمون خارج هذه العملية وأحيانًا لا يكونون على دراية بمشاكل التعلم التي يواجهونها في فصولهم. يمكن أن يكون الحل هو اعتماد منهجية تحليل المشاعر لمراجعات الطلاب من أجل اكتشاف الحالة المزاجية والقضايا أثناء عملية التعلم. سيضمن ذلك حل المشكلات وسرية الاتصالات بين الطلاب. وبالتالي ، مما يؤثر بشكل مفيد على تجربة الطلاب ككل في منصات التعلم الإلكتروني

أظهرنا في هذه الرسالة محاولة استخدام تحليل المشاعر في مراجعات التعلم الإلكتروني لاكتشاف مشاكل الطلاب في التعلم الإلكتروني. ركزنا بشكل خاص على اقتراح طريقة التغذية الراجعة التي تضمن سرية الاتصال وتساعد على تحسين تجربة التعلم الإلكتروني.

الكلمات المفتاحية: التعلم الإلكتروني، تحليل المشاعر، تحليل الرأي، التعلم الآلي، معالجة البيانات، ترشيح البيانات.

Acknowledgement

First of all, we would like to thank the jury members for kindly participating in the evaluation of this work.

Also, we would like to express our sincere thanks to all the Saad DAHLAB University teachers and administrative staff of computer science department, who contributed, directly or indirectly from near or far to the realization of this work. Their lessons, comments, suggestions and advices affected our academic and life path.

With sincere feeling of respect and gratitude we thank our supervisor, Mrs. Melyara Mezzi, we are grateful for the trust deposited in our work and for the motivation demonstrated along this arduous period. Her support, advice, guidance and patience during the preparation of this thesis was without a doubt crucial.

Special thanks to our families, friends and classmates their support, encouragement and love gave us the strength to make this work. They, undoubtedly continue to inspire us and allows us to see the future as an endless challenge.

Finally, we thank ALLAH for giving us the power, the patience, the will to finish this work during this hard period.

Chahrazed and Safia

Dedications

This study is wholeheartedly dedicated my beloved parents, who have been my source of inspiration and gave me strength when I thought of giving up, who continually provided their moral, spiritual, emotional, and financial support.

I dedicate my dissertation work to my family and many friends. A special feeling of gratitude to my loving parents, their words of encouragement and push for tenacity ring in my ears. My brothers Mohamed and Farah, have never left my side and are very special.

I dedicate this work and give special thanks to my best friend Safia Yami and her family also for being always there for me and for their love, I am so blessed to have them in my life, and I would to thank my friend Affaf feroukhi for her input. They have been my best cheerleaders.

I would to thank warmly my supervisor Mrs. Melyara Mezzi, my teacher for accepting the responsibility to supervise us, for her love and care, thank you for all that you did, for me you are a star and I love you so much, we are lucky to have you in our life

Through this thesis, I pay special tribute to my grandmother, I miss you incredibly much. Even though you are no longer with me, I can see your smile on your face and I know you are watching over and showering your blessings to grace the moment.

And lastly, I dedicated this thesis to almighty god, thank you for guidance, strength, power of mind, protection and skills and for giving us a healthy life. All of these we offer to you.

Chahrazed DJIDA

Dedications

First and foremost, I would like to dedicate my work to my parents and my sibling: Mounia, Lilya and Sid-Ahmed, and my grandma. Who have been my source of inspiration and endless support, who continually provided their help and never left my side.

I sincerely thank all my teachers and classmates who have played an important role in my academic path and personal development.

I would like to thank my supervisor Mrs. Melyara Mezzi, for the guidance, direction, and pearls of wisdom, but more importantly, for putting up patiently with mails and questions while providing amazingly encouraging feedback, and offering encouragement precisely when needed and without which it would have been nearly impossible to produce this piece of work during this hard time.

Special thanks to my best friend Djida Chahrazed, who shows me every day that no matter how different we are and what happens in this world some people will always find a way to be there for each other. My Friends: Affaf, Selma, Sophia, Oualid, Sofiene and Mohamed for endless support and advises. Thank you for being part of my life.

Through this thesis, I pay special tribute to my grandfathers, as you can see I'm still trying not to be a mess.

Finally, Allah who gave me the power, the patience, the will to finish this work during this hard period.

Safia YAMI

Table of contents

General introduction	18
1. Research background	3
2. Research problems	4
3. Research challenges	4
4. Thesis outline	4
Chapter I: State of the art about E-Learning	6
1. Introduction.....	7
2. The evolution of E-Learning.....	7
2.1. The iMooX MOOC Platform.....	11
3. Learning methods.....	13
4. The Relevance of E-Learning in Higher Education.....	14
4.1. In Algeria	14
4.2. In the world.....	15
5. The advantages and disadvantage of E-learning.....	16
5.1. The advantages of E-learning.....	16
5.2. The disadvantages of E-learning	18
6. E-Learning systems	19
6.1. Architecture of e-learning systems.....	19
6.2. Communication process	21
6.3. The actors in e-learning	23
5. Criteria for an E-learning Platform	24
6. Conclusion	26
Chapter II : Sentiment Analysis.....	27
1. Introduction.....	28

2.	Definition.....	28
3.	The evolution of Sentiment Analysis	29
4.	Level of analysis	30
5.	Classification of sentiment analysis approaches.....	31
5.1.	Machine Learning Approach	32
5.2.	Lexicon-based Approach.....	35
5.3.	Hybrid Approach.....	35
6.	Related works.....	37
6.1.	Sentiment Analysis for Social Media	37
6.2.	Evaluation of E-Learning activity effectiveness in higher education through sentiment analysis	38
6.3.	A Sentiment Analysis System to Improve Teaching and Learning	38
6.4.	Sentiment Analysis in MOOC Discussion Forums.....	39
7.	Advantages of using sentiment analysis.....	41
8.	Disadvantage of using sentiment analysis	41
9.	Challenges.....	41
10.	Conclusion.....	43
	Chapter IV: Conception and solution modeling	44
1.	Introduction.....	45
2.	Software architecture.....	45
3.	Data preprocessing	47
3.1.	Data cleaning	48
3.2.	Data Transformation	48
3.3.	Data Reduction.....	48
4.	Sentiment analysis.....	49

4.1. Rule-based approach	49
4.2. Automatic Sentiment Analysis	50
5. N-Gram:.....	50
6. Filtration	51
7. Project Life Cycle	52
6.1. Verification phases.....	54
6.2. Validation phase.....	54
8. Requirement specification and analysis	54
7.1. Identification of the actors	54
7.2. Use Case diagram of the system	55
7.3. Sequence Diagram	57
9. System design	60
10. Conclusion.....	62
Chapter V: Implementation of the solution.....	63
1. Introduction.....	64
2. Tools and languages used	64
3. Architecture of the solution	65
3.1. Datasets.....	65
3.2. Database.....	67
3.3. Data preprocessing and filtration	68
3.4. Sentiment analysis.....	72
4. Presentation of the platform.....	74
4.1. Home	75
4.2. Courses	75
4.3. Login	76

5. Signup.....	83
6. Test and evaluation.....	83
6.1. Test part.....	83
6.1.1. Evaluation part	84
7. Conclusion.....	88
General conclusion	89
1. Summary.....	90
2. Outlook.....	90
References	91

List of figures

Figure 1: MooX MOOC platform homepage.....	12
Figure 2: Video conferencing System Architecture (9).....	15
Figure 3: Learning Models types.....	20
Figure 4: Connection between learning systems' parts.	21
Figure 5: Sentiment Analysis process on product reviews (16).....	29
Figure 6: Approaches of Sentiment Analysis.....	32
Figure 7: Machine learning classifier.	33
Figure 8: General architecture of the proposed solution.....	46
Figure 9: General process of the system.	47
Figure 10: The V-Shaped Software Development Model.....	53
Figure 11: Use case diagram of ELearning platform.....	56
Figure 12: Sequence diagram of the student's interactions with the platform.....	59
Figure 13: Sequence diagram of the teacher's interactions with the platform.....	60
Figure 14: Class diagram of the platform.	61
Figure 15: Learness home page.....	75
Figure 16 : Learness courses page.....	76
Figure 17: Learness login page.....	76
Figure 18: Administrator profiles page.....	77
Figure 19: Administrator reviews page.....	77
Figure 20: Administrator uploading reviews.....	78
Figure 21: Admin checking preprocessing results.....	79
Figure 22: Admin checking polarity results.	79
Figure 23: Admin checking filtration results	80
Figure 24: Teacher adding new course.....	80

Figure 25: Teacher checking the feedback (unfinished).....	81
Figure 26: Student checking the course table.....	81
Figure 27: Student checking the forum.....	82
Figure 28: Student writing a review.	82
Figure 29: SignUp page.	83
Figure 30: Checking the feedback as teacher.....	84

List of tables

Table 1: Some Popular Online Education Initiatives.....	10
Table 2: Communication components in e-learning.....	22
Table 3: Sentiment classification approaches.....	36
Table 4: Related works about the use of Sentiment Analysis in E-Learning.....	39
Table 5: ELearning platform service and functionalities.	57
Table 6: Class diagram explanation.	61
Table 7: Problem-solution dataset description.	66
Table 8: Database description.	67
Table 9: Regular expressions examples.	70
Table 10: Words polarity and type according to textblob.....	73
Table 11: Detections performed by the system VS Detections performed by human.....	86
Table 12: Negative-neutral review detection evaluation.	86
Table 13: Problems detection evaluation.	87

Acronyms list

Acronym	Definition
----------------	-------------------

MOOC:	Massive Open Online Course
--------------	----------------------------

NLP :	Natural Language Processing
--------------	-----------------------------

NLTK:	Natural Language Toolkit
--------------	--------------------------

SA:	Sentiment Analysis
------------	--------------------

UML:	Unified Modeling Language
-------------	---------------------------

General introduction

1. Research background

Our society is living an impressive transformation, maybe the most important during the past couple of years, which through the strong diffusion of the new technologies, is radically modifying the nature of the relationships among countries, markets, people and cultures. This technological revolution has clearly facilitated the process of globalization and the information exchange. In this way, information can be considered as an economic good whose value is tightly linked to the amount of knowledge that can be given to its users. Gaining new knowledge, competences or skills has determined the need for a continuous update by the actors of the supply chain of the new economy.....

E-Learning represents an effective answer to the continuous request of life-long learning. Many institutions are adopting this approach both to improve their traditional courses and increase the potential audience since it allows more flexibility and quality in general. E-classrooms are often composed by students inattentive or appearing bored and wondered. This sense of frustration increases when it faces students' poor performance on tests. Recent studies showed that emotions can affect the e-learning experience. A general definition for emotions is the following: emotions are complex psychophysical processes that evoke positive or negative psychological responses (or both) and physical expressions, often involuntary.

Working towards improving e-learning experience, it is important to know students opinions about the course and course tools. Course forums provide students with the chance to engage in social learning and share their issues in e-learning platforms. Analyzing the data from this part of the course, can allow the inference of important information about what can be improved in the course.

2. Research problems

The topic of emotions goes far beyond this simple definition and it is especially hard to detect in an E-Learning environment because students feedback is needed. These reviews and comments shared usually in forums and chat rooms are often outside of teachers' sight so they do not understand the learning problems and issues encountered in their classes. Another problem is the large quantity of reviews that makes it impossible to read them all.

3. Research challenges

The main objective of this thesis is to automate the process of problem detection from student reviews and send it to the teacher. This automatization will ensure the confidentiality of communication and allow huge amount of data to be processed.

4. Thesis outline

The contribution of this work is an investigation into how sentiment analysis of student's reviews can improve the e-learning experience. The reminder of this thesis is structured as follows:

- Chapter 1 will contain a state of the art about E-learning. In this chapter, we will present the evolution of E-learning. Then, we will talk about E-learning methods, followed by the relevance of E-learning in higher education. Finally, we will present some E-learning systems and highlight the most important criteria of E-learning platforms.
- Chapter 2 concerns the notions behind Sentiment Analysis. In this chapter, we will introduce the evolution of Sentiment Analysis together with the level of analysis and the classification of Sentiment Analysis approaches. Then, we will present some related works and expose the advantages, disadvantages, and the challenges of using

Sentiment Analysis techniques in Natural Language Processing (NLP) applications.

- Chapter 3 will cover the modeling of the solution we propose so as to improve students' experience in E-learning platforms. We will first present our software architecture and then give a detailed explanation of each component of the solution using a V-shaped software development model based on some UML diagrams to clarify both the requirements and functionalities to be in the final application. Finally, we will present the software design presented in the form of a class diagram which we will explain minutely.
- Chapter 4 focus on the technical and practical aspects of the solutions. In this latter, we will first present the tools and languages that have been used. We will then talk about our dataset, database, and give some back-end execution examples of the essential steps of the solution. Finally, we will set forth our demonstration platform through its screenshot and its evaluation.
- Finally, the genera conclusion of this thesis will summarize the theoretical and practical work that have been done to the aim of finding a solution to the enhancement of students' experience in E-learning platforms and settle some outlooks and future works.

Chapter I:
State of the art about
E-Learning

1. Introduction

E-learning has witnessed a great interest from corporations, educational institutions and individuals alike. As an education pattern, e-learning systems have become more and more popular. It commonly refers to teaching efforts propagated through the use of computers in a bid to impart knowledge in a non-traditional classroom environment. As a prerequisite for an effective development of e-learning systems, it is important to have certain knowledge about users' opinions and build an evaluation regarding them [1].

E-Learning represents an effective answer to the continuous request of life-long learning. Many institutions are adopting this approach both to improve their traditional courses and increase the potential audience since it allows more flexibility and quality in general.

The first chapter of this thesis opens with the background of the study and further describes the field of E-Learning. We're taking things back to basics to introduce the uninitiated to the wonderful world of E-Learning. We're addressing why we might want to consider adding E-Learning as a method of course delivery and how the right tools can be brought together.

2. The evolution of E-Learning

Formally, E-learning can be defined as follows: *“E-learning is learning utilizing electronic technologies to access educational curriculum outside of a traditional classroom. In most cases, it refers to a course, program or degree delivered completely online [2]”*.

The growth of e-learning is more of a dynamic process as it is affected by many factors such as the subject matters, the knowledge level of the audience and the environment around the audience. The users also play an important role in how they want to learn as all users do not have same approach to learn. A major factor that affects the E-learning process is the amount of information that the user wants to receive. The users must decide

for themselves which information is important for them at a moment of time and how much they can process. These factors have vastly affected the way E-learning has developed from the early times to the present.

During the early 1980's when computers were scarcely available instructors-led training was the only way of learning things from or about computers. Computers were only available for official use, and if staffs wanted to learn from computers they had to use the office hours. Using the office hours to train led to major slowdown in the official work and was rather counterproductive for the offices.

Then arrived the era of multimedia advancements from around 1985 to 1995 Windows 3.1¹ .Macintosh², CD ROMs and PowerPoint started making their marks. Then to make training portable and visually engaging, instructors started using CD ROMs to deliver instructions. As the instruction could be easily transported using a CD and could be delivered repeatedly even without the presence of any instructor, this era represented a major upgrade from instructor led training. This also resulted in major cost savings and was more productive compared to an instructor led training method. The major downside of this training method was the lack of the instructor's interactions and dynamic presentations, which left the audience wanting more.

During the middle 1990's when the World Wide Web (www) started evolving, instructors started exploring the possibilities of using the new technology to deliver instruction in more efficient and cost effective ways. This period saw the growth of email, web browsers, HTML³, media players, low definition audio/video files and basic Java applications. With the growth of these technologies, the way in which instruction was delivered had also evolved. Instructors and students started communicating via emails and the

¹ **Windows 3.1:** is a series of 16-bit operating environments produced by Microsoft for use on personal computers, released on April 6, 1992.

² Macintosh: A family of personal computers from Apple and the first to popularize the graphical user interface (GUI).

³ HTML: HTML is an acronym for Hypertext Markup Language which is defined as the coding used to include color, language, and graphics on a webpage.

use of audio/video files in delivering instruction started to emerge. The major downfall was the size of the instruction files. The larger files could not be delivered via the web and it was very costly to use the web during this period. To deliver via storage devices was also not possible because the floppy discs were the commonly used storage devices then and had a really low storage capacity (720KB)

Then in the early 20th century the world started seeing a major growth in the field of computers and webs. People started moving from the age of CD ROMs to USB and low bandwidth web to the high bandwidth web. Faster and portable computers were replacing slower and bigger computers. The Java/IP network applications, high definition audio/video files started surfacing

All these changes started to change the way people learn. The instructors started to have real-time conversation with students on the web and bigger files were easily transferred via USB⁴ or were easily accessible via the web. When slowly smartphones and hand-held devices with the Internet started gaining popularity, it completely changed the way people looked for information. The term *e-learning* also started to gain attention [3].

With the growing popularity of hand-held devices the effectiveness of e-learning has grown even further. With use of the Internet in hand held-devices people were able to access instructions or information with just a single click. Today the e-learning process has come a long way from the instructor-led training era to a hand-held era.

Table 1 lists some popular online education initiatives. Although these initiatives are booming exponentially, the debate remains about whether corporate-based online-education initiatives have the credentials to grant

⁴ USB: USB is defined as an acronym that stands for Universal Serial Bus which is a type of computer port which can be used to connect equipment to a computer.

credits and degrees to individual learners who have completed their online courses [4].

Table 1: Some Popular Online Education Initiatives.

Name	Sponsor	Date	Fees
TED-Ed	Sapling Foundation	1984	Private Non-profit
MITx	Joint efforts by Harvard University and edX	2001	Non-profit
Khan Academy	Salman Khan (Hedge Fund manager)	2007	Non-profit
Saylor	Michael J. Saylor (Chairman, CEO, & President of the business intelligence company MicroStrategy)	2008	Non-profit
University of the People	Shai Reshef (educational entrepreneur)	2009	Non-profit
Peer 2 Peer University (P2PU)	Funding from the Hewlett Foundation & the Shuttle worth Foundation	2009	Non-profit
Udemy	Eren Bali	About 2010	Partly free
eduMOOC	University of Illinois Springfield	2011	Free
Coursera	Joint efforts by Princeton	2011	Private

	University, Stanford University, University of California Berkeley, University of Michigan-Ann Arbor, & University of Pennsylvania		
edX	Harvard University & MIT	2012	Non-profit
iTunes U	Apple Corporation	2012	Non-profit
Minerva	Minerva project and Keck Graduate Institute (KIG). (Larry Summers, former Harvard University President & United States Secretary of the Treasury, chaired its first advisory board)	2012	private
Udacity	Sebastian Thrun	2012	For-profit

2.1. The iMooX MOOC Platform

Since education in the countries of Central Europe is primarily offered as face-to-face classes, the need for distance learning was a must, especially for a country like Austria [5] As long as MOOCs (Massive Open Online Courses) provide an excellent opportunity for online learners and a strong basis of distance education, this paves the way to encourage decision makers to adopt a project to establish a MOOC platform in Austria. iMooX is an online learning stage and the first Austrian xMOOC platform. It was founded in 2013 as a result of collaboration between the University of Graz and Graz University of Technology. The platform went online in February 2014. And the figure 1 shows MooX MOOC platform homepage.

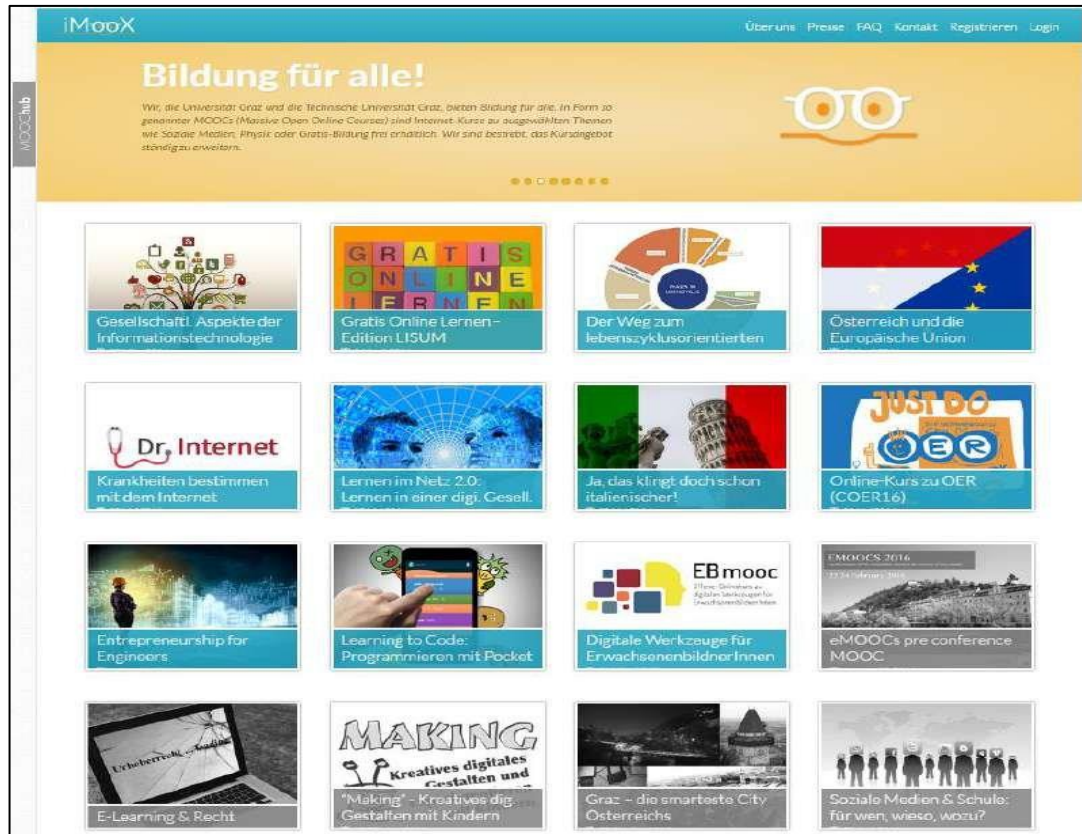


Figure 1: MooX MOOC platform homepage.

The design of the iMooX platform endeavors to follow the cognitive-behaviorist pedagogy theme concepts of Gagne⁵ (1965) [6]:

- Acquiring the learners' attention by providing them the correct steps to gain the learning theory through the online education system.
- Listing the objectives and learning goals of each online course.
- Demonstrate the stimulus by presenting active online learning videos.
- Giving feedback through discussion forums and regular emails
- Assessing performance through computerized assessment of the exams.
- Providing guidance, which usually depends on the learners themselves where self-learning is imperative due to the online learning environment conditions.

⁵ Robert Mills Gagne: was an American educational psychologist best known for his Learning conditions.

3. Learning methods

Learning methods are referred to as ways through which instructors deliver instructions and learners access these instructions. Several learning methods have been described in literature, including traditional learning, ELearning, blended learning, mobile learning, and personalized learning [7].

- **Traditional learning** refers to face-to-face sessions, through which the teacher delivers course material to students in the same place and at the same time. The learning method is teacher centered, where the teacher focuses on providing the learning information to the students. Assessments depend on study notes given to students by the teacher, limiting the learner's knowledge acquisition boundary.
- **E-Learning**, on the other hand, refers to the use of ICTs to transform and support the learning process ubiquitously. For instance, Meyen et al. [8] define e-Learning as the acquisition and use of knowledge which is distributed and facilitated primarily by electronic means. Such electronic means may include internet, intranet, extranet, CD-ROM, video tape, DVD, TV, and personal organizers often adopted.
- **Blended learning** makes use of a combination of various learning methods that include face-to-face classroom activities, live e-Learning, and self-paced learning. This learning method encompasses a variety of tools for simulating and maximizing the learner's learning potential. This has resulted in the adoption of this learning method in various higher education institutions of learning.
- **Mobile learning** is defined as learning or delivery of content that is facilitated by the use of portable technologies such as mobile phone, PDAs (Personal Digital Assistants), or iPods .The global penetration and the use of mobile technologies have created new avenues and enhancements in teaching and learning activities in higher education Currently, mobile learning presents vast benefits that facilitate and enhance e-Learning.

- **Personalized learning** is a learning approach that facilitates and supports individualized learning. Each learner has a learning path that caters for learners learning needs and interests in a productive and meaningful way. One of the attributes of personalized learning is the ability to dictate the students' learning.

Advancements in technology have led to a paradigm shift from traditional to personalized learning methods with varied implementation strategies. In the following, we will explore the e-Learning state-of-the-art over the past decade.

4. The Relevance of E-Learning in Higher Education

4.1. In Algeria

To compensate for the lack of supervision on the one hand, but also in order to improve the quality of training, in accordance with the requirements of quality assurance, other forms of learning and teaching, incorporating new educational approaches in the training process are being implemented in Algeria. This is how the national system of distance education was launched, and whose objectives are divided into three steps [9]:

- The first phase of technology uses video conferencing in particular, to absorb the flow of learners, while significantly improving the quality of teaching and training (short-term processes).
- A second step, which uses new educational technologies, based especially on the Web (online learning or e-learning), to achieve quality assurance (medium-term process).
- An integration phase in which the system of distance education, is deployed and validated, to ensure distant teaching, with the key, creating a chain of knowledge, whose use and benefit will far outweigh the only academic world, to reach a wider audience

of learners who want to expand their knowledge, people with special needs, senior citizens, hospital patients, people undergoing rehabilitation, etc.

Currently, the National System of Distant Learning Network is based on a platform of video conferencing and e-learning, scattered throughout almost all training institutions. Access to this network is provided by the National Research Network (ARN)⁶. And Figure 2 shows video-conferencing System Architecture.

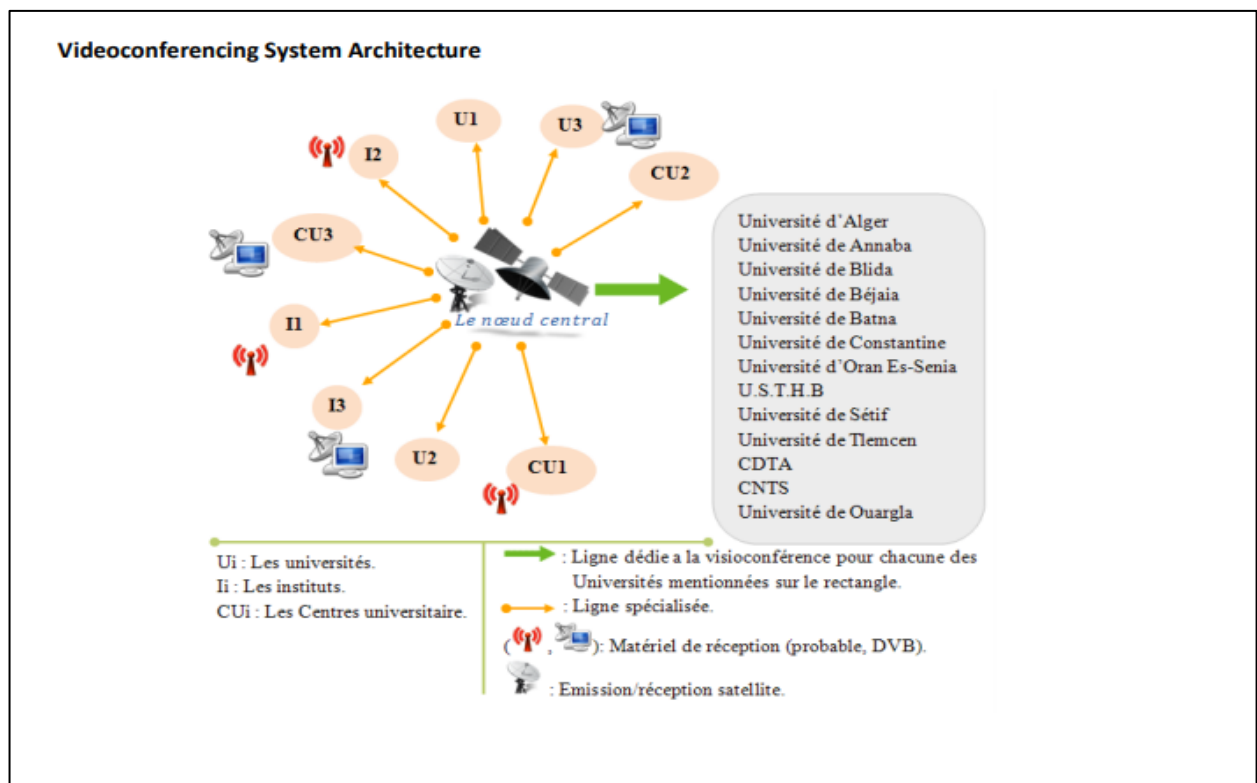


Figure 2: Video conferencing System Architecture [9].

4.2. In the world

Students can use E-learning for attending e-classes from the comfort of their home, whilst at the same time being able to see teachers and their

⁶ ARN: The ARN network provides national and international connectivity and evolves regularly with the evolution of technology and the capacity of the available infrastructures.

classmates, receive and submit homework and exams, exchange class notes and files. It is easy to learn and easy to apply the solution to any situation that bars a student from physically attending school.

Let's take a look at how COVID-19 has affected education and the ways that student information systems can provide useful tools for an uninterrupted educational experience. But they do not lose the ability to continue their education from the comfort and safety of their own homes. Student Information Systems offer unique e-learning capabilities that could resolve physical presence barriers for any reason, but also in cases of emergency situations such as the outbreak of the Coronavirus. Students are attending their classes normally from a distance by signing on to an E-class. E-classes offer much more than just convenience. They offer easy note-taking, recording of the classes to listen to it later, easy exchange of learning material and homework. A student is capable of doing all the things that would be able to do in a physical classroom excluding the horseplay [10].

5. The advantages and disadvantage of E-learning

Hereafter, we will expose the advantages and drawbacks of E-learning practices.

5.1. The advantages of E-learning

E-Learning has many advantages. With the combination of a well-organized e-learning system and a highly motivated student, one can achieve great success in a short period of time. Some of the major advantages of e-learning are listed below [11]:

- Convenient for students: E-learning materials are self-placed and can be accessed any time the learner wants. They do not require the learner to be physically present in a classroom. Students can also download and save the learning materials for future purposes from the system.

- Lower cost: E-learning is usually a cost-efficient way of learning for most students as they can choose from a large range of courses and make the selection depending on their needs. It can also be cost-efficient for many universities because once the learning platforms are set up, they can be reused for many sessions.
- Up-to-date learning materials: The study materials in e-learning systems can be updated more frequently than in the classroom-based education systems. Once the study materials are placed in the system, they can be updated without changing the whole materials and the materials can be available and reused for longer times.
- Flexible way of learning: E-learning is a flexible way of learning for many students. Most of the study materials are stored for the students to access whenever they want. Students can also choose between an instructor-led and a self-learning system. In E-Learning systems it is also possible for students to skip over the study materials they already know and choose to focus on the ones they want to learn.
- World-wide learning society: E-Learning systems help in creating a worldwide learning society as anyone can access the study materials regardless of the geographical location. In the systems available now, learners can also contribute to the study materials, which help to keep the materials updated.
- Scalable e-learning systems: The number of students in virtual classes or e-learning systems can be very few or really high without causing any significant difference in the total cost.
- Higher degree of freedom for students: One may find it difficult at times to learn new ideas. E-Learning systems provide the possibility for students to learn the same material repeatedly until they are satisfied.
- Better retention: The video and audio materials used in e-learning make the whole learning process more fun. This will help students to remember the things they learn for a long period of time. E-learning

materials can also be accessed whenever wanted, thus the repetition makes the retention easier.

5.2. The disadvantages of E-learning

As any other systems e-learning also have some drawbacks. Being flexible is not always good as it may cause laziness and thus reduce efficiency. Some of the major disadvantages of E-learning include [11]:

- Low motivation: Students with low motivation may not achieve the set goals most of the time, as there is no one to look over. Students are themselves responsible for the routine and organization of the course, thus leading to laziness and low motivation at times. Lack of a fixed schedule and deadlines may lead to students dropping out of the course prematurely.
- Technology-dependency: The study materials in e-learning are delivered using computer applications. For some people just to learn how to use those applications might take a long period of time. There are also other factors such as a poor Internet connection and machine malfunction, which may make the learning process tedious and time consuming.
- Compatibility issues: As there are many learning systems available, sometimes the study materials prepared by using one system may not be compatible with another. For example mobile devices such as iPads block the flash videos from being played in their browsers. Access to websites may not be free in all places as some countries impose restrictions on the number of websites that can be accessed.
- Reliability of the content: The content available on the Internet might not always be reliable. There are people who mislead the readers and feed wrong information. So the readers must be careful while searching for the information and check the reliability of the content before learning it.

- Social isolation: Lack of a real classroom or classmates might not be good for all students. Students might feel socially isolated at times due to lack of real people around while learning.
- Expenses management: In the long run e-learning is usually a cheaper option but still for the first time it might prove too expensive for some institutions. Buying new equipment like computers, projectors or new software all at once might not always be easy for new or small institutions. Also the cost of developing training materials is high compared to the traditional methods.
- Disadvantages for disabled students: While developing new study materials or study environments companies usually target at large user groups. They might not always think about disabled students such as visually impaired students as they might not gain from it as the development costs may be high.
- Not effective in all cases: In some cases, face-to-face study materials might be more effective than learning online as e-learning lacks two-way communication.

6. E-Learning systems

6.1. Architecture of e-learning systems

The main requirement of the developers while developing any e-learning system is that it should be easy to use and safe. Students should be able to choose their course, find the resources easily, track their progress and reach their final goal on time. In the learning system students should be able to communicate and help each other, in other words it should be social and engaging.

While developing the learning system the designer should focus on creating a perfect mix of traditional and virtual learning methods to help the students engage more and reach their goals.

A perfect e-learning system is a mix of all the learning systems shown in figure 3. In theory the perfect e-learning system will allow students to learn from their home while having some real-time support and interaction with other students who are taking the same courses however it might be difficult to achieve in real cases. The types of E-Learning models are shown in figure 3.

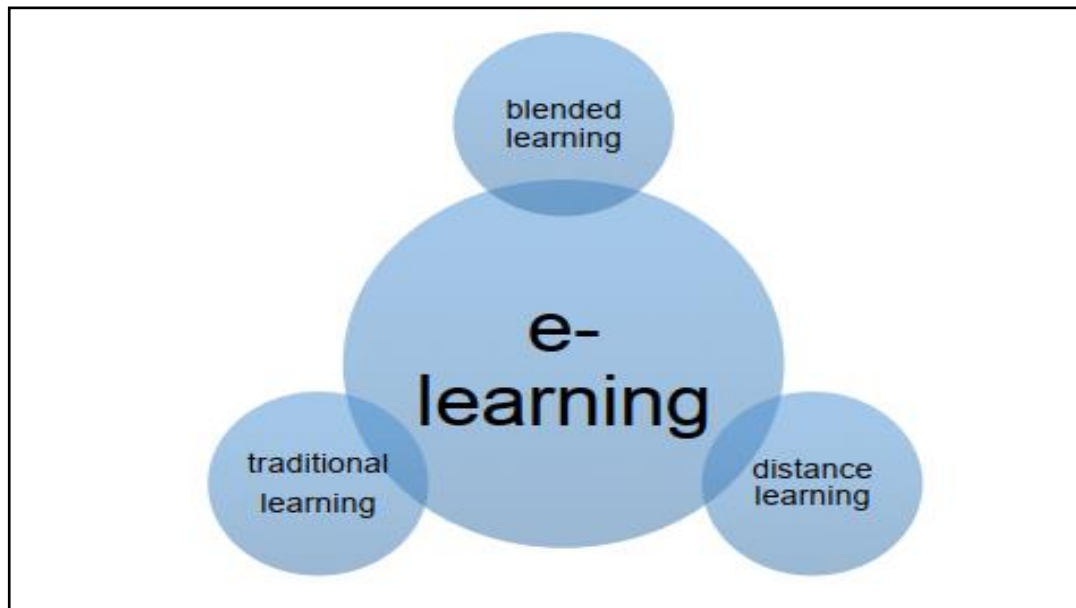


Figure 3: Learning Models types.

The architecture of an e-learning system might be difficult to understand in technical term but it usually consists of three parts:

1. User Interface.
2. Platform.
3. Database.

A *database* is the place where the digital learning materials are stored and a *learning platform* is the media from which the students access the learning materials. A *user interface* is what the student sees while accessing the platforms. The communication process between these three parts is very important for the whole system to work perfectly. For example, we can use MySQL to store the digital materials, use edX platform to access the materials and students can visit the edX platform using their web browser.

The connection between the three parts of the e-learning system is shown in figure 4.

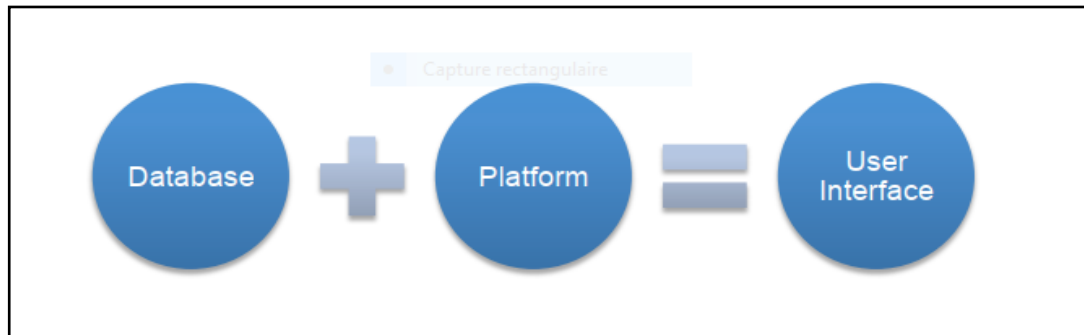


Figure 4: Connection between learning systems' parts.

As we can see from the figure 4, the user interface is what the student sees while accessing the platform, thus giving access to the database in the background.

- **The user interface:** is the means by which the student interacts with the system, including the input devices and the software. It allows user to input their command through input devices such as a mouse or touch screen and shows the effects of the user input. Applications and web browsers are some examples of the user interfaces.
- **The database:** is a structured set of information or data organized so that it can be accessed, managed and updated whenever necessary. In order to store, modify, extract and administer the database a management system, database management system (DBMS)⁷ is required. There are many database management systems available in the market like Oracle, Microsoft SQL Server, and MySQL [11].

6.2. Communication process

In the whole e-learning system, the communication process plays a very important role. Whether it is during the preparing phase i.e. between the people involved during the creation phase or during the training phase as

⁷ DBMS: Software that controls the organization, storage, retrieval, security and integrity of data in a database. It accepts requests from the application and instructs the operating system to transfer the appropriate data.

we saw in the architecture above, the communication process is necessary. The lack of proper communication in any phase of the system may not produce the desired result or even cause the whole system to crash. But the process of communication during the development and the training phase may differ [11].

People involved during the modeling phase are the ones who prepare the materials and the ones who make those digital copies available to the users.

But between that, there is a host of other people involved such as illustrators, drawers, recorders, camerapersons and photographers who have to work together to produce the necessary learning materials.

After the study material has been prepared, they will be stored as a digital copy in the database. Users access the materials form the database using different platforms. The communication process used generally here is the network.

In the whole communication process the technology used is the hardware, software and the network communication. But creators, host and learners use these components in different levels. In general the communication process can be summarized as in Table 2.

Table 2: Communication components in e-learning.

Communication component	Creators	Host	Learner
Hardware	Multimedia devices to create the materials, example: cameras	Network equipment	Computer or hand held devices with network connections
Software	Necessary applications to create the material, example: Microsoft	Servers	Web browsers, flash and media players

Office			
Network connection speed required	Moderate speed	High speed	Moderate speed

Table 2 illustrates the different components required in the communication process during the different phases of the e-learning system.

6.3. The actors in e-learning

Famous philanthropist and great tycoon Andrew Carnegie once said: *"Teamwork is the ability to work together toward a common vision and the ability to orient individual achievements toward organizational goals. It is the fuel that enables ordinary people to achieve extraordinary results"*. Close collaboration, a strong work ethic, and shared goals are the hallmarks of an effective eLearning project. To achieve this, however, the E-learning team must be chosen wisely. Each professional must play their role in the process in order to offer an eLearning course that exceeds expectations and achieves educational objectives, on time, within budget.

Here are the 4 key players to have on a shock eLearning team [12, 9]:

1. **The Administrator:** The administrator refers to the author who deals with the software maintenance. Indeed, he ensures the management of the accounts of the other actors playing a role in the e-learning. Also, he maintains in good working order, the learning system which integrates new information and communication technologies.
2. **The Course Creator:** The course creator has the role of course development through using the platform tools, according to the teaching objectives. Also, he introduces changes according to the learners' and the tutors' reactions. Indeed, he facilitates learning by

managing the actors and the events in order to ensure the good progress of the learning process.

3. The Learner: The learner can be defined as a student or an employee in a learning situation which uses a platform to acquire information and to transform them into knowledge. Indeed, the learner is evaluated. He prepares reports, projects, and takes part in discussion forums.

In turn, the authors have focused on the role of the trainers which proves to be insufficient to overshadow certain individual characteristics of the learners in a learning environment. Indeed, they refer to several researchers to present these characteristics. It is about: the motivation, the fear or frustration, and the desire to learn and add that the learners shall develop an active behavior, interact with the contents, adapt the contents, get committed, and be autonomous for a successful e-learning.

4. The Tutor: The tutor is part of a pedagogical support activity for the learner. The authors define the tutor as an intervention for one or more learners aiming at their personal and social development, through inviting him to assume his responsibilities in relation to his own learning. In addition, the tutor's mission is the monitoring and the supervision of the learners throughout their e-learning. Thus, the tutor provides assistance to the learners. Indeed, they argue that supervision is a fundamental dimension of learning to the extent that the tutor provides educational assistance to the learner corrects the work and provides him feedback.

5. Criteria for an E-learning Platform

According to the "Guide for designing and developing e-learning courses" [13], a learning platform is a set of interactive online services that provide learners with access to information, tools and resources to support educational delivery and management through the Internet. Usually, there are 3 kinds of learning platforms [14]:

1. Virtual learning environments (VLEs)
2. Learning management systems (LMSs)
3. Learning content management systems (LCMSs).

These definitions have no clear limits and are often used interchangeably. There are certainly differences between them but some of the features of these platforms are common.

1. Virtual learning environments (VLE): These are learning platforms used to simulate traditional face-to-face classroom activities and facilitate teaching and learning. Their main characteristic is their strong collaborative component. The most well-known VLEs are “Moodle” and “Blackboard”.
2. Learning management system (LMS): Using this kind of learning platforms, we can facilitate the delivery and management of all learning offerings, including online, virtual classroom and instructor led courses. It also automates the learning course, delivers easily the training, manages learners and keeps track of their progress and performance across training activities, which reduces administrative overhead. The differences between VLE and LMS comes more from the setting in which they operate. LMSs are, primarily, for training while VLEs are primarily for education. The well-known Moodle platform is considered to be an LMS within corporate e-learning, but it is referred to as a VLE in the education sector where it promotes a communicative and collaborative approach
3. The 3rd type of learning platform, the “learning content management system” focuses mainly on creating e-learning content. So, it is used mainly by developers and administrators to create content material for e-learning and blended learning courses. This material includes articles, tests, games, video and small units of digital content, called content chunks.

6. Conclusion

There is a strong need for identifying suitable strategies for effective e-Learning implementation and we have here provided a general overview of various theories for learning processes and methods. We have analyzed some quite recent E-Learning implementation trends and discussed e-Learning implementation aspects.

In the next chapter we will focus on the study of the Sentiment Analysis.

Chapter II :

Sentiment Analysis

1. Introduction

With the emergence of World Wide Web during the last two decades, internet has become the leading source of information. This source of information contains enormous amount of human generated reviews on products, services, government policies, social issues, etc. on different blogs, social media, chat forums, manufacturer's or distributor's websites. These reviews not only help people to make a choice for buying some specific product, but also give a vibrant idea to manufacturers about their products or services.

For any customer or manufacturer, it is almost impossible to read and analyze properly all these reviews manually and build a decision upon them. Which made need for its efficient computer processing emerged very quickly. Sentiment analysis deals with the extraction of sentiments from user generated content (UGC) helping us to understand the social sentiment of their brand, product or service.

In this chapter, we will talk about the notion of sentiment analysis, its evolution together with levels of analysis, the classification of its approaches and finally some related works and challenges.

2. Definition

Sentiment Analysis [15] refers to the application of natural language processing, computational linguistics, and text analytics to identify and classify subjective opinions in source materials (document or a sentence).

Generally, sentiment analysis classifies text expressions in source materials into two types: *facts* (objective expressions about entities) and *opinions* (subjective expressions of sentiments, attitudes, emotions or feelings toward entities). It should be pointed out that not all subjective sentences contain opinions and not all objective sentences contain no opinions.

Therefore, it is important for sentiment analysis to identify and extract facts and opinions from source text materials. More generally, the basic tasks of sentiment analysis mainly includes: *sentiment identification* (subjectivity identification) and *sentiment orientation classification* (determine the orientation of an opinionated text). Figure 5 hereafter, shows Sentiment Analysis process on product reviews.

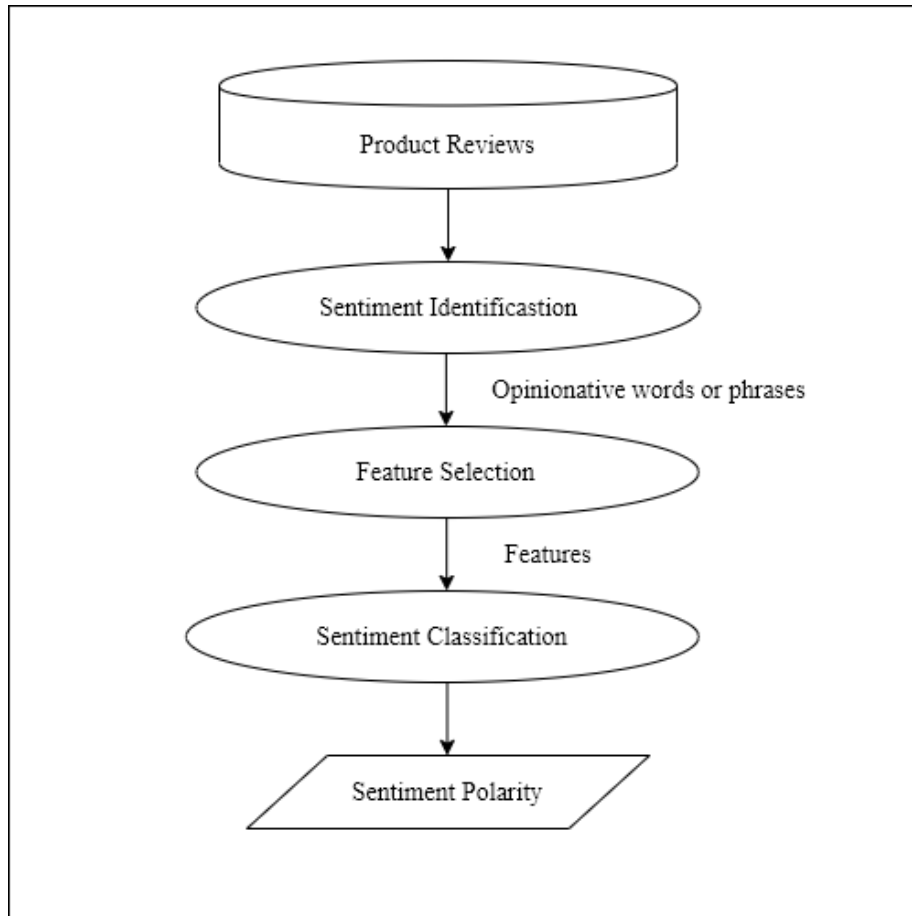


Figure 5: Sentiment Analysis process on product reviews [16].

3. The evolution of Sentiment Analysis

It is difficult to pinpoint when sentiment analysis first appeared. Reviews and feedbacks have always been considered valuable for companies to know how public see their brand and services.

Certainly by the early 2000s, as more and more businesses began to maintain an online presence, there was thought being put into the idea.

Via email and online sales, businesses started asking customers for feedback on the product and their experience interacting with the company. These types of surveys were expensive, labor intensive and very limiting. This method relied heavily on those with an interest in the industry and who liked to participate in a survey or give any feedback.

As online presence grew, it became clear that technology was lagging behind the need for good sentiment analysis. The explosion of social media moved the industry and the concept of sentiment analysis in the right direction, but it still did not go all the way.

Suddenly businesses could gauge public perception of their brands in new ways: the amount of likes on their social media accounts, the ratings, and reviews on their products and the ability to search for articles and blogs about them. While these all provided valuable new tools for companies, it was soon discovered that this too was lacking.

Then along came machine learning and advanced algorithms. The technology was finally catching up. In just the last few years, self-improving algorithms have taken over sentiment analysis operations, automatically scanning the internet and supplying valuable data on the overall positivity or negativity of public sentiment.

Sentiment analysis is now at the forefront of technological horizons, taking the path of advanced AI and versatile customization and evolution. It is taken some time to develop, but the future of sentiment analysis looks like a bright one.

4. Level of analysis

In general, sentiment analysis has been investigated mainly at three levels [17, 18]:

1. **Document level:** analysis resolves whether a positive or negative sentiment predominates about subject to which the document refers.

But it is not applicable to documents which evaluate or compare multiple entities.

2. **Sentence level:** the sentence can be viewed as a small document, and therefore there is no fundamental difference between the sentiment analysis at the document level and at the sentence level.
3. **Aspect level:** both the document level and the sentence level analyses do not discover what exactly people liked and did not like. Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion). An opinion without its target being identified is of limited use. Realizing the importance of opinion targets also helps us understand the sentiment analysis problem better (example reviews of restaurants like in the sentence 'Food is excellent, but the service is terribly slow', in here aspects are food and services). The goal of this level of analysis is to discover sentiments on entities and/or their aspects.

5. Classification of sentiment analysis approaches

According to [16, 17, 18, 19] classification of sentiment analysis approaches can be divided into three approaches as summarized in figure 6:

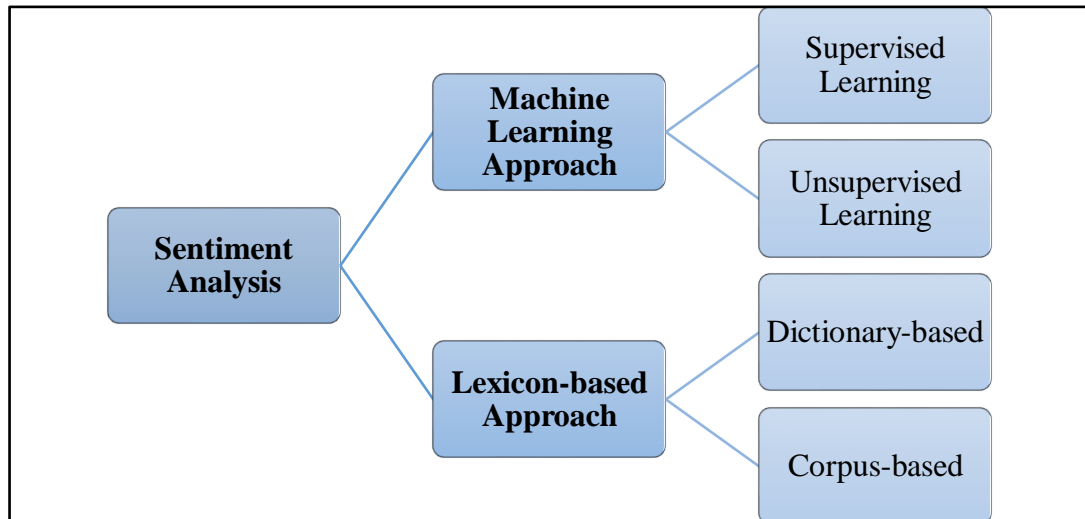


Figure 6: Approaches of Sentiment Analysis.

5.1. Machine Learning Approach

A sentiment analysis task using machine learning approach is usually modeled as a classification problem, whereby a classifier is fed a text and returns a category, e.g. positive, negative, or neutral. It can be divided into supervised and unsupervised learning methods. The supervised methods make use of a large number of labeled training documents. The unsupervised methods are used when it is difficult to find these labeled training documents.

And it consists of two sets of documents: training and a test set. The training set is used for learning the differentiating characteristics of a document, while the test set is used for checking how well the classifier performs. Below, is how a machine learning classifier can be implemented (Figure 7):

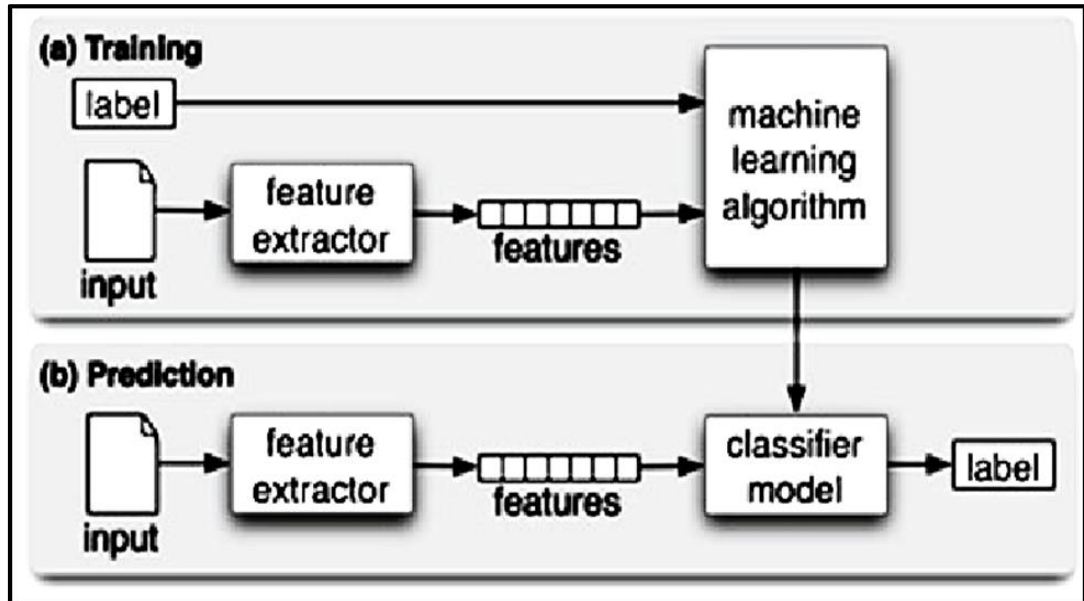


Figure 7: Machine learning classifier.

The above figure shows how generally a classifier is being trained and then predicts labels for unknown inputs. Identifying a better feature extraction method is the most important thing in building a classifier.

5.1.1. The Training and Prediction Processes

In the training process (a), the model learns to associate a particular input (i.e. a text) to the corresponding output (tag) based on the test samples used for training. The feature extractor transfers the text input into a feature vector. Pairs of feature vectors and tags (e.g. positive, negative, or neutral) are fed into the machine learning algorithm to generate a model.

In the prediction process (b), the feature extractor is used to transform unseen text inputs into feature vectors. These feature vectors are then fed into the model, which generates predicted tags (again, positive, negative, or neutral).

5.1.2. Feature Extraction from Text

The first step in a machine learning text classifier is to transform the text extraction or text vectorization, and the classical approach has been bag-of-words or bag-of-ngrams with their frequency.

More recently, new feature extraction techniques have been applied based on word embedding (also known as word vectors). This kind of representations makes it possible for words with similar meaning to have a similar representation, thus improving the performance of classifiers.

5.1.3. Supervised classification most popular algorithms

The classification step usually involves a statistical model and the most popular are:

- **Naïve Bayes (NB):** A simple probabilistic classifier that uses Bayes Theorem to predict the category of a text.
- **Maximum Entropy (ME):** A probabilistic model that estimates the conditional distribution of the class label.
- **Support Vector Machines (SVM):** A representation of the examples as points in space in which support vectors are computed to provide a best division of points/examples into categories.
- **Logistic Regression Model (LR):** A LR model predicts the classes from a set of variables that may be continuous, discrete or a mixture.

The features of machine learning based approach for sentiment classification are:

- **Term presence and their frequency:** that includes uni-grams or n-grams⁸ and their presence or frequency.
- **Part of speech information:** used for disambiguating sense which is used to guide feature selection.
- **Negations:** has the potential of reversing sentiments opinion words/phrases: that expresses positive or negative sentiments.

⁸ In the fields of computational linguistics and probability, an n-gram is a contiguous sequence of n items from a given sample of text or speech. In this regard, a uni-gram is an n-gram consisting of a single item from a sequence

5.2. Lexicon-based Approach

The lexicon based approach uses sentiment dictionary with opinion words and match them with the data for determining polarity. There are three techniques to construct a sentiment lexicon: manual construction, corpus-based methods and dictionary-based methods.

- The manual construction is a difficult and time-consuming task.
- Corpus-based methods can produce opinion words with relatively high accuracy.
- Dictionary based techniques, the idea is to first collect a small set of opinion words manually with known orientations, and then to grow this set by searching in the WordNet⁹ dictionary for their synonyms and antonyms.

5.3. Hybrid Approach

Hybrid systems combine the desirable elements of the machine learning and the lexicon based approaches into one system. It has the potential to improve the sentiment classification performance providing more accurate results.

There are some advantages and limitations in using these different approaches depending on the purpose of the analysis. The main advantage of machine learning approaches is the ability to adapt and create trained models for specific purposes and contexts, while the limitation is that it is difficult integrating into a classifier, general knowledge which may not be acquired from training data. Furthermore, learnt models often have poor adaptability between domains or different text genres because they often rely on domain specific features from their training data.

Lexicon-based approaches have the advantage that general knowledge sentiment lexicons have wider term coverage, however these approaches

⁹ WordNet is a lexical database of semantic relations between words in more than 200 languages. WordNet links words into semantic relations including synonyms, hyponyms, and meronyms.

have two main limitations. Firstly, the number of words in the lexicons is finite, which may constitute a problem when extracting sentiment from very dynamic environments. Secondly, sentiment lexicons tend to assign a fixed sentiment orientation and score to words, irrespective of how these words are used in a text.

The main advantages of hybrid approaches are the lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain. While the main limitation is that reviews are with a lot of noise (irrelevant words for the subject of the review) are often assigned a neutral score because the method fails to detect any sentiment.

Table 3 summarizes the reviewed techniques of this section:

Table 3: Sentiment classification approaches.

Approaches	FEATURES/TECNIQUES	ADVANTAGES AND LIMITATIONS
Machine Learning	<ul style="list-style-type: none"> - Term presence and frequency - Part of speech information - Negations - Opinion words and phrases 	<p>Advantages:</p> <ul style="list-style-type: none"> - The ability to adapt and create - Trained models for specific purposes and contexts. <p>Limitations:</p> <ul style="list-style-type: none"> - The low applicability to new data
Lexicon-based	<ul style="list-style-type: none"> - Manual - Corpus-based - Dictionary based 	<p>Advantages:</p> <ul style="list-style-type: none"> - Wider term coverage. <p>Limitations:</p> <ul style="list-style-type: none"> - Finite number of words in the lexicons and the assignation of a fixed sentiment orientation and score to words.

Hybrid	<ul style="list-style-type: none"> - Sentiment lexicon constructed using public resources for initial sentiment detection - Sentiment words as features in Machine Learning method 	<p>Advantages:</p> <ul style="list-style-type: none"> - Lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain. <p>Limitations:</p> <ul style="list-style-type: none"> - Noisy reviews
--------	--	---

6. Related works

6.1. Sentiment Analysis for Social Media

It is a very important fact to analyze how people think in different context about different things. This becomes more important when it comes to the business world. Promising results has shown in [19] that through sentiment analysis in social media can be used to cater business environment needs.

The data were gathered from social media Twitter¹⁰ to extract sentiments out of them in order to be used later in product profiling, trend analysis and forecasting. Then use lexical data source SentiWordNet¹¹ that gives positive and negative score for each word, to have polarity value. Even though, sometimes it did not give correct polarity values for sentences it was overall capable of saying that how a set of people of a particular age range, a particular area with a particular profession think about a particular product or service and how it will change in the future which are most useful information when it comes to business world.

¹⁰ Twitter is an American microblogging and social networking service on which users post and interact with messages known as "tweets."

¹¹ SentiWordNet is document resource which contains a list of English terms which have been attributed a score of positivity and negativity.

6.2. Evaluation of E-Learning activity effectiveness in higher education through sentiment analysis

Sentiment Analysis has been a part of text mining research domain and recognized due to its successful implementation in social media analysis. The research [20] took place, E-Learning's users opinion were used as object to evaluate the E-Learning activity. Implemented in Widyatama University in Indonesia which had been running E-Learning activities for several years.

Qualitative method by given questioner to users and gather the feedback were commonly used as evaluation of E-Learning systems. Still, questioner doesn't represent the conclusion about the whole opinion.

The opinion datasets have been labeled into positive and negative which divided into training and testing set. Naïve Bayes algorithm has been used as learning method and shows the accuracy by 87.5%. Another evaluation was also performed with precision, recall and F-measure to represent relevant and irrelevant document and they showed good results.

6.3. A Sentiment Analysis System to Improve Teaching and Learning

Many universities obtain such feedback via a Student Response System (SRS) during or at the end of a course to analyze the teacher's performance. Another sentiment analysis system [21] used student feedback about teacher performance, the learning experience, and other course attributes can also be gathered through social media.

This system analyzed student surveys and comments obtained from a university SRS system for 25 different courses over a two-year period and compared the percentage of positive sentiments, students had about each course with the average course grade on a 0–100 scale.

Despite its good performance, the system has some limitations. It is only as good as the data it analyzes, so care must be taken in collecting feedback from students. SRSs must be well designed to ensure that they are engaging, and instructors must make a concerted effort to ensure that as many students as possible provide complete and accurate feedback.

6.4. Sentiment Analysis in MOOC Discussion Forums

In order to monitor student’s trending opinions towards the course and major course tools. Sentiment Analysis is used to study drop out behavior in Massive Open Online Course (MOOC) [22]. It explores the relation between opinions expressed by students and the students’ dropout rate by extracting the positive and negative sentiment words that are most associated with the course tool topic keywords.

In order to understand the impact of sentiment on the user-level, it examines how sentiment that members have expressed and are exposed to in a particular week, predicted their continued participation in the forum discussion. This analysis showed that sentiment analysis should be used with caution in practice, especially when the texts are very noisy and limited in quantity.

Table 4 summarizes the above mentioned works:

Table 4: Related works about the use of Sentiment Analysis in E-Learning.

Research work	Purpose	Dataset	Results/observation
Sentiment Analysis for Social Media [19].	Generate knowledge that can be useful for business environments using people’s attitudes about their products and	Twitter dataset.	Sometimes it did not give correct polarity values but overall it was capable of saying how a set of people feel about a particular product or service.

	services on social media.		
Evaluation of E-Learning activity effectiveness in higher education through sentiment analysis [20].	Gather E-Learning participant opinion in questioner and see if it would lead to proper advice for decision making process.	E-Learning system at Widyatama University.	It showed an accuracy of 87.5% and the evaluation performed with precision, recall and F-measure showed good results too.
A Sentiment Analysis System to Improve Teaching and Learning [21].	Use students' feedback, course grade and other course attributes to improve the process.	Coursera dataset.	Despite the good performance, the system has some limitations. The SRSs must be well designed and instructors must ensure that students provide complete and accurate feedback.
Sentiment Analysis in MOOC Discussion Forums [22].	Find the relation between opinions expressed and the students' dropout rate.	Coursera dataset.	There is a correlation between sentiment expressed in the course forum posts and the number of students who drop the course. But it should be used with caution.

7. Advantages of using sentiment analysis

Among the advantages of sentiment analysis, we can cite:

- Sentiment Analysis makes it easier to gauge how users feel by analyzing their comment without having to read thousands of them at once.
- Having thousands or even tens of thousands of survey responses per month, makes it impossible for one person to read all of these responses and have an unbiased and consistent measure of user's sentiment. By using Sentiment Analysis and automating this process, it is made easier to drill down into different segments of user's comments and get a better understanding of sentiment in these segments.

8. Disadvantage of using sentiment analysis

Even though Sentiment Analysis is a useful tool, it cannot be a complete replacement for reading survey responses, because often there are useful nuances in the comments themselves. But sentiment analysis can be helpful further by identifying which of the comments should be read.

9. Challenges

Given the fact that the sentiment analysis is carried out over an unstructured text full of context, the success of the analysis depends on a number of factors. Some of the obstacles to successful analysis are the distinction between facts and opinions, syntax errors, negations, comparative sentences, irony and sarcasm recognition, fake opinions, context dependency, domain dependency, lack of resources for non-English languages, relatively small number of annotated datasets for supervised machine learning and the unsuccessfulness of the model in the domain where it was not developed.

Detecting subjectivity is not trivial to humans because it is a subjective task in itself - the sentence is for someone neutral or not neutral because of different levels of expertise on the topic or a different interpretation of the sentence.

Social network users are writing more freely and often do not care about writing proper words. Sometimes they deliberately write the words incorrectly to point out the thought (for example, they write the word baaaad, whose meanings are very bad) or they replace the letters by number (2day instead of today). Shortening is common (e.g. whatews instead of whatever), and it is not unusual for the two short words to join in a new one (e.g. crunk from crazy / drunk). In addition, the text can be filled with jargon and slang.

Negations are an almost unavoidable challenge in the sentiment analysis because they most often change the polarity of opinion.

A special problem is irony and sarcasm that even people cannot always understand, and this is a big problem for the computer during the analysis.

The increase in the number of fake opinions is also one of the major problems. Researchers try to find ways of detecting such reviews as they can significantly affect the outcome. One of them is the analysis of unusually positive interactions of potential spammers with ordinary users in order to make their opinion more influential.

Depending on the context, one and the same word can be both positive and negative and also neutral. For example, the word “long” is basically neutral “long hair” but becomes negative in the example of “long queue” and positive in “long battery”. The word context is often far-off the word it depends on in the sentence, so it is difficult to understand when looking at the phrases in which words stand next to each other.

Due to hard and time consuming work, a relatively small number of labeled data sets required for supervised machine learning are evident.

Most of the research in this area is focused on texts written in English, and consequently the largest number of resources (labeled data sets, dictionaries ...) is developed for that language. To overcome this issue, researchers perform automatic translations in which part of the text semantics is lost and the success of the analysis is reduced.

10. Conclusion

Given all the above-mentioned facts, it is obvious that sentiment analysis is becoming more and more important and valuable nowadays.

It has not been fully implemented in E-Learning sector yet. But in order to make improvement in teaching and learning process it should be used in that sector too.

In the next chapter, we will present the modeling of our modest contribution to meet the project's objectives.

Chapter IV:
Conception and solution modeling

1. Introduction

Software development plays an important part in modern computing world. And the software development models are the foundation for software engineering. The goal of a software process model is to provide guidance for systematic coordinating and controlling of the tasks that must be performed in order to achieve the end product and the project objectives. For software development process, a process model defines the following: a set of tasks that need to be performed, input to and output from each task, preconditions and post conditions for each task, and the sequence and flow of these tasks.

The Unified Modeling Language (UML) is a standard visual modeling language that allows to specify, visualize, and document models of software systems [23]. UML is used to show the whole architecture of an application by proposing diagrams, all different from each other but linked, helping the understanding of the system.

In this chapter, we will give a short description of the conception and modeling tools, and then talk about the different tools that we have used to develop our solution.

2. Software architecture

Software architecture functions as a blueprint for the system and the developing project, laying out the tasks necessary to be executed. Software Architecture diagram is a crucial step to describe the basic software structure. It depicts how a typical software system might interact with its users, external systems, data sources, and the figure below (Figure 8) shows the architecture of our Automatic Sentiment Analysis (ASA) program in E-learning platform.

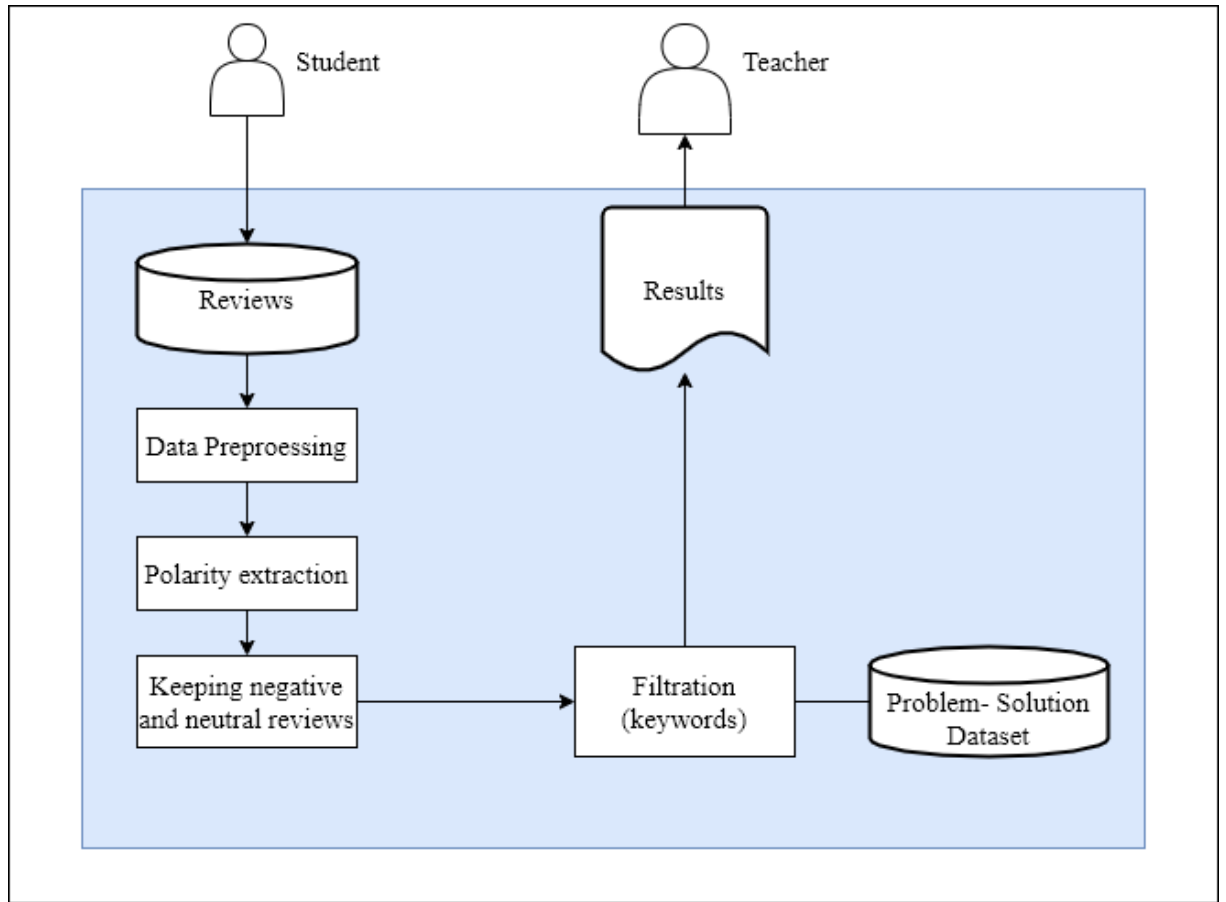


Figure 8: General architecture of the proposed solution.

As it can be seen the figure above, the general process of our solution can be broken up to six steps:

- **Step 1: Getting the reviews**
Getting students' reviews about the course.
- **Step 2: Data preprocessing**
Processing the reviews with regular expressions.
- **Step 3: extracting the polarity**
Use sentiment analysis functions to define the polarity of the review.
- **Step 4: keeping the negative and neutral reviews**
After the polarity extraction, we will keep only the negative and neutral reviews to work with.
- **Step 5: Filtration**

By using keywords of the reviews and problems, we can filter and detect the problems.

- **Step 6: sending the results**

After gathering all the results we can send the problems located and their succinct solutions to the teacher.

The figure below (Figure 9), summarizes the above mentioned process:

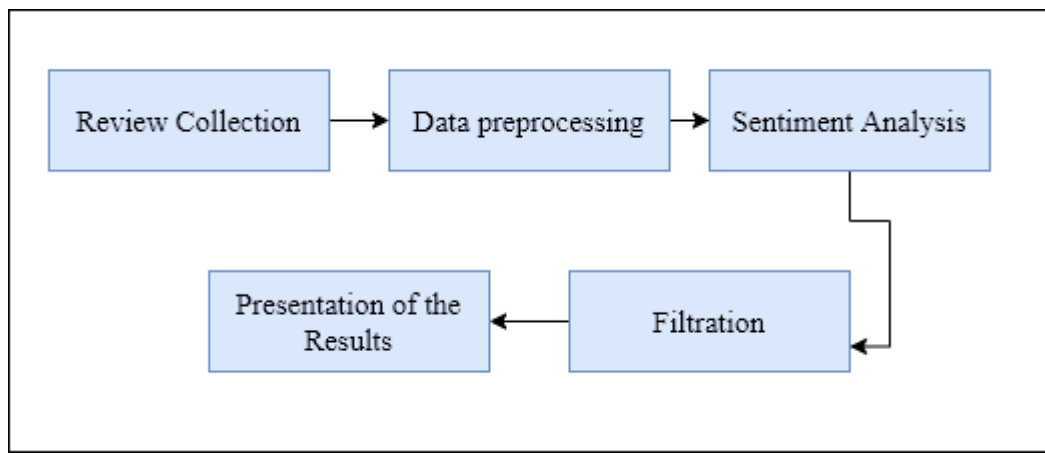


Figure 9: General process of the system.

In the reminder of this chapter, we will give a point by point explanation of the different components of the solution, but for now, some theoretical basis about data processing, sentiment analysis and filtration steps are needed and will be presented hereafter.

3. Data preprocessing

Data preparation, known also as data pre-processing, is focused mainly on two tasks: firstly, the data must be organized into a proper form, and, secondly, the datasets used must lead to the best performance and quality for the models obtained by data mining operations [24, 25].

Real-world data tend to be incomplete, noisy and inconsistent. This can lead to a poor quality of collected data, and, further to a low quality of models built on such data. Reducing data dimensionality by identifying and

removing as much irrelevant and redundant data as possible, allows building patterns faster and more effectively.

In order to address these issues, data preprocessing contains operations which perform data cleaning, data transformation and data reduction. The most important objectives of data reduction are: to avoid over fitting and improve model performance.

3.1. Data cleaning

The raw data can have incomplete records, noise values, outliers and inconsistent data. Data cleaning is the first step in data preprocessing which is used to find the missing values, smooth noise data, recognize outliers and correct inconsistency. These dirty data can lead to unreliable and poor output. Therefore, it is important to do some data cleaning.

3.2. Data Transformation

This step is taken in order to transform the data in appropriate forms suitable for mining process.

3.3. Data Reduction

These techniques are used to reduce the representation of dataset in smaller volume with respect to maintain the integrity of the original dataset. And while working with huge volume of data, analysis became harder in such cases. In order to get rid of this, we uses data reduction to increase the storage efficiency and reduce data storage and analysis costs.

The goal of data preprocessing is to obtain high-quality data for any type of mining. Once we are done with the data pre-processing tasks, we can carry on to sentiment analysis.

4. Sentiment analysis

As explained in sentiment analysis section, sentiment analysis is used for analyzing attitudes, opinions, views and emotions from text, speech, tweets and database sources in the publicly available social networks.

In the previous section we showed that sentiment analysis techniques can be divided into machine learning approach, lexicon based approach and hybrid approach. And in this section we will see major Sentiment Analysis methods that we might implement.

4.1. Rule-based approach

Rule-based sentiment analysis is based on an algorithm with a clearly defined description of an opinion to be detected. Includes identify subjectivity, polarity, or the subject of opinion.

The rule-based approach involves basic Natural Language Processing routine. More precisely, it involves the following operations upon the text corpus:

- Stemming.
- Tokenization.
- Part of speech tagging.
- Parsing.
- Lexicon analysis (depending on the relevant context).

There are two lists of words. One of them includes only the positive ones, the other includes negatives. The algorithm goes through the text, finds the words that match the criteria. After that, the algorithm calculates which type of words is more prevalent in the text. If there are more positive words, then the text is deemed to have a positive polarity.

The thing with rule-based algorithms is that while it delivers some sort of results - it lacks flexibility and precision that would make them truly usable.

4.2. Automatic Sentiment Analysis

Instead of clearly defined rules - this type of sentiment analysis uses machine learning to figure out the gist of the message. And it involves classification algorithms such as:

- Linear Regression.
- Naïve Bayes.
- Support Vector Machines.
- RNN derivatives LSTM and GRU.

Below is the pseudo code of our polarity review algorithm:

Algorithm: polarity extraction (review):

```
SW = {Set of words and their polarity}
nb = { number of reviews words}
Input: review
Output: polarity score
Begin
  For every word of a review do
    sum = word polarity according to SW
    nb= nb+1
  End for
  score = sum/ nb
  return score
End
```

After the sentiment analysis task, comes the filtration phase.

5. N-Gram:

An n-gram defines a subsequence of n items from a given sequence. It is used in various fields of natural language processing and genetic sequence analysis. An n-gram model defines a method for finding a set of n-gram words from a given document. The commonly used models include unigrams (n=1), bigrams (n=2) and trigrams (n=3). However the value of n can be extended to higher level grams. The n-gram model can be better explained with the following examples:

Text: "Honesty is the best policy."

Unigrams: "honesty", "is", "the", "best", "policy".

Bigrams: "honesty is", "is the", "the best", "best policy".

Trigrams: "honesty is the", "is the best", "the best policy".

Unigrams presents the simplest model for the n-gram approach. It consists of all the individual words present in the text. The bigram model defines a pair of adjacent words. Each pair of words forms a single bigram. The higher order grams can be formed in the similar way by taking together the n adjacent words. Higher order n-grams are more efficient in capturing the context as they provide better understanding of the word position [26]

6. Filtration

Filtration is the most important phase of our system, since we have to detect eLearning problems from our dataset in the reviews.

And in recent years, semantic similarity measure has had a great interest in Semantic Web and Natural Language Processing. Several similarity measures have been developed, being given the existence of a structured knowledge representation offered by ontologies and corpus which enable semantic interpretation of terms.

Semantic similarity between concepts is a method to measure the semantic similarity, or the semantic distance between two concepts according to a given ontology. In other terms, semantic similarity is used to identify concepts having common "characteristics" [27].

The problem with this method is that it takes in consideration only the closest concept. So if a review talks about two or three eLearning problems, we will not be able to find them all. For this we have to find another method that will detect all the eLearning problems in the review.

And now that we established the architecture of our platform and the processes, we can move to the modeling and conception section.

7. Project Life Cycle

The "software lifecycle" [28, 29] refers to all stages of software development, from the initial idea to its fruition. The objective of such a breakdown is to make it possible to define intermediate milestones allowing the validation of software development, that is to say the conformity of the software with the expressed needs and the verification of the development process i.e. the adequacy of the methods used.

The origin of this division comes from the observation that errors have a higher cost when they are detected late in the production process. The life cycle makes it possible to detect errors as early as possible and thus to control the quality of the software, the deadlines for its production and the associated costs.

The life cycle model adopted for the realization of this project is the V-model, since our project is medium sized where requirements are clearly defined and fixed, the figure below (Figure 10) presents this model in its different stages:

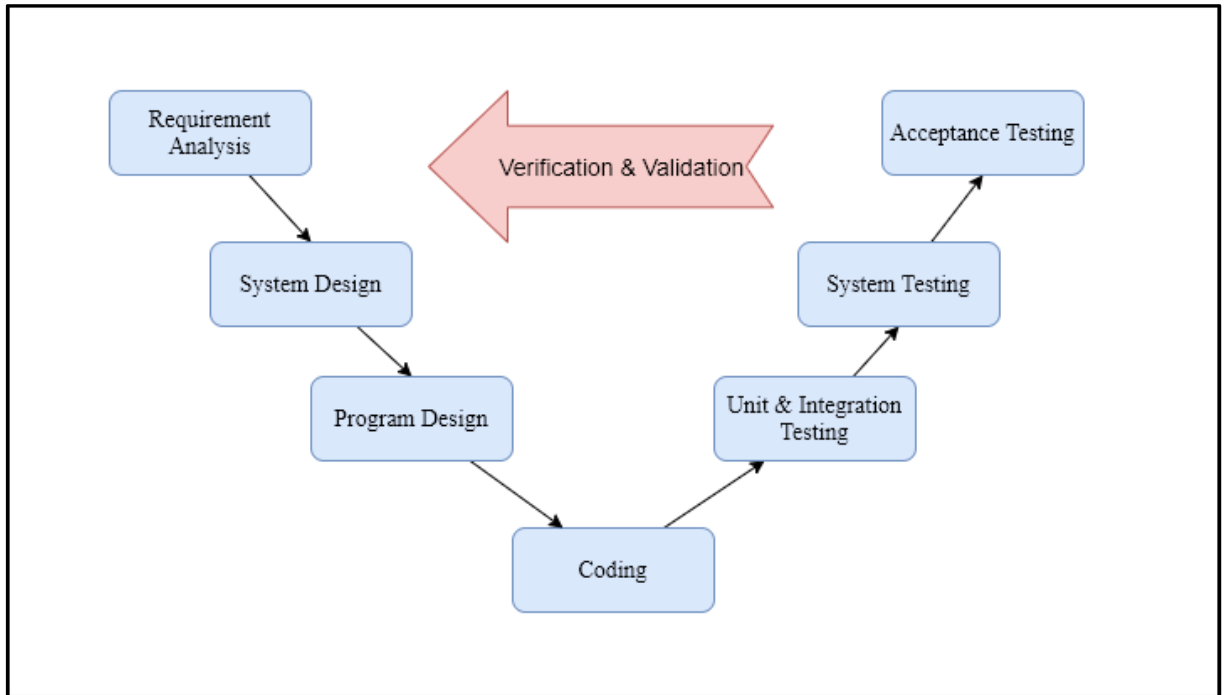


Figure 10: The V-Shaped Software Development Model.

The V-model means Verification and Validation model [28]. It is considered as an extension of the waterfall model. Instead of moving down in a linear way, the process steps are bent upwards after the coding phase, to form the typical V shape.

This model comprises of four main process phases: requirements, specification, design and implementation and has a corresponding verification and validation testing phase. Implementation of modules is tested by unit testing, system design is tested by Integration testing, system specifications are tested by system testing and finally, acceptance testing verifies the requirements. The V-model got its name from the timing of the phases. Starting from the requirements, the system is developed one phase at a time until the lowest phase, the implementation phase, is finished.

At this stage testing begins, starting from unit testing and moving up one test level at a time until the acceptance testing phase is completed. During development stage the program will be tested at all levels simultaneously.

The different levels of the V-Model are: unit tests, integration tests, system tests and acceptance test. The unit tests and integration tests ensure that the system design is followed in the code. The system and acceptance tests ensure that the system does what the customer wants it to do. The test levels are planned so that each level tests different aspects of the program and so that the testing levels are independent of each other.

6.1. Verification phases

It involves static analysis technique (review) done without executing code. It is the process of evaluation of the product development phase to find whether specified requirements meet.

6.2. Validation phase

It involves dynamic analysis technique (functional, non-functional), testing done by executing code. Validation is the process to evaluate the software after the completion of the development phase to determine whether software meets the customer expectations and requirements.

8. Requirement specification and analysis

The requirement specification and analysis is an important step in the software development. This phase makes it possible to better clarify the actors, Functional and non-Functional needs and other requirements expected in order to have a better understanding of the subject.

7.1. Identification of the actors

This step is one of the most important. It consists on understanding the context of the system. This involves determining the functionalities and the actors.

In UML, we will use the term actor to designate the people who interact with the platform. An actor is an external entity which makes it

possible to identify all that the platform has to offer to the actor through different roles that the latter must play to highlight all the functionalities.

- **The admin:**

The admin is a unique player in our case, he will be able to connect to an interface which will show him the different interactions of teachers and students with the platform (like creation of new course or new review). He can then consult the courses, users (Teachers and Students) and even the reviews.

- **The teacher:**

Using a specific identifier, the teacher will be able to access the platform management course interface. The teacher must be able to add, modify and delete the course. He will also receive the problems found in his course and their solutions.

- **The student:**

The student is an important player in the E-Learning platform, for this he must access to a simple interface which must group all the expected features such as check courses, take course, check forum and write a review.

7.2. Use Case diagram of the system

The use case allows have a clear view of all the requirements of our solution. Use cases specify the expected behavior (what?), and not the exact method of making it happen (how?). A key concept of use case modeling is that it helps us design a system from the end user's perspective. The diagram bellow (Figure 11), represents the use case diagram of our E-Learning platform.

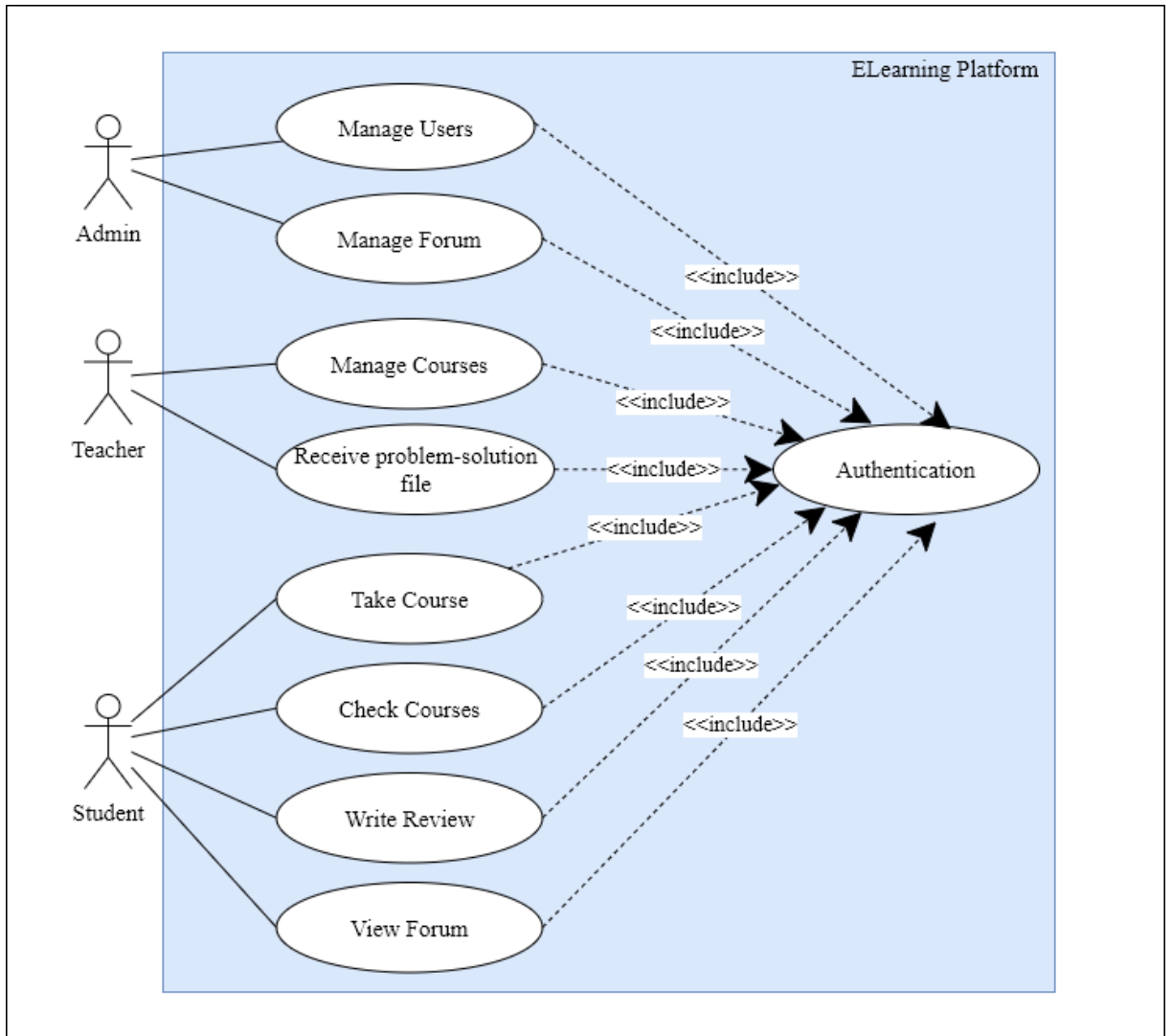


Figure 11: Use case diagram of ELearning platform.

This use case diagram exposes the different functionalities of our ELearning platform. And since our project is more focused on improving learning process with students reviews we are not going to focus on authentication part. Table (Table 5) below gives more details.

Table 5: ELearning platform service and functionalities.

Actor	Service function	Description
Admin	Manage users	The admin can edit or delete users.
	Manage reviews	The admin can view or delete review on forum.
Teacher	Manage course	The teacher can create, edit or delete a course.
	Receive problem-solution file	The teacher can receive problem-solution file about his course.
Student	Check course	The student can check on the existing courses.
	Take course	The student can take a course.
	Write review	The student can write a review about a course.
	View reviews	The student can view other reviews about a course.

7.3. Sequence Diagram

Sequence diagram is one of the common interaction diagrams used in UML [23], which focuses on the message interchange between a number of object lifelines. Sequence diagram describes an interaction by describing the sequence of messages that are exchanged, and also explains their corresponding occurrence specifications on the lifelines.

Since our platform is more focused on the problem detecting and improvement of learning process we will focus on 2 main scenarios:

- a. The student taking the course and writing the review.
- b. The teacher creating the course and getting the problem-solution file.

7.3.1. Scenario 1: The student's interactions

The scenario of the student taking a course and writing the review is detailed below:

- *Actors*: The actor who can take a course and write the review is the student.
- *Pre-conditions*: Authentication.
- *Starting action*: Access the platform.
- *Chain of actions*:
 - Authentication (verification of user name and password).
 - Access to the platform.
 - Selection and assessing the course.
 - Writing a review about the taken course.
- *End action*: exit the platform.
- *Post-conditions*: None

The diagram below (Figure 12) illustrates the sequence diagram of this scenario.

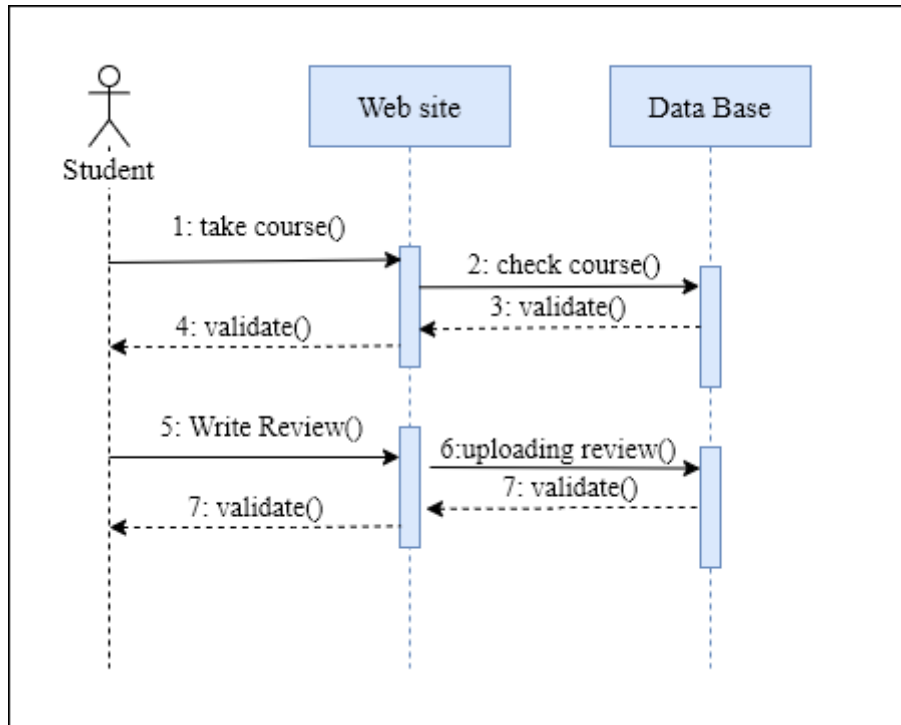


Figure 12: Sequence diagram of the student’s interactions with the platform.

7.3.2. Scenario 2: The teacher’s interactions

The scenario of the teacher creating the course and getting the problem-solution file is detailed below:

- *Actors:* The actor who can create a course and get the problem-solution file is the teacher.
- *Pre-conditions:* Authentication.
- *Starting action:* Access the platform.
- *Chain of actions:*
 - Authentication (verification of user name and password).
 - Access to the platform.
 - Creating the course.
 - Receiving the problem-solution file.
- *End action:* exit the platform.
- *Post-conditions:* None

The diagram below (Figure 13) illustrates the sequence diagram of this scenario.

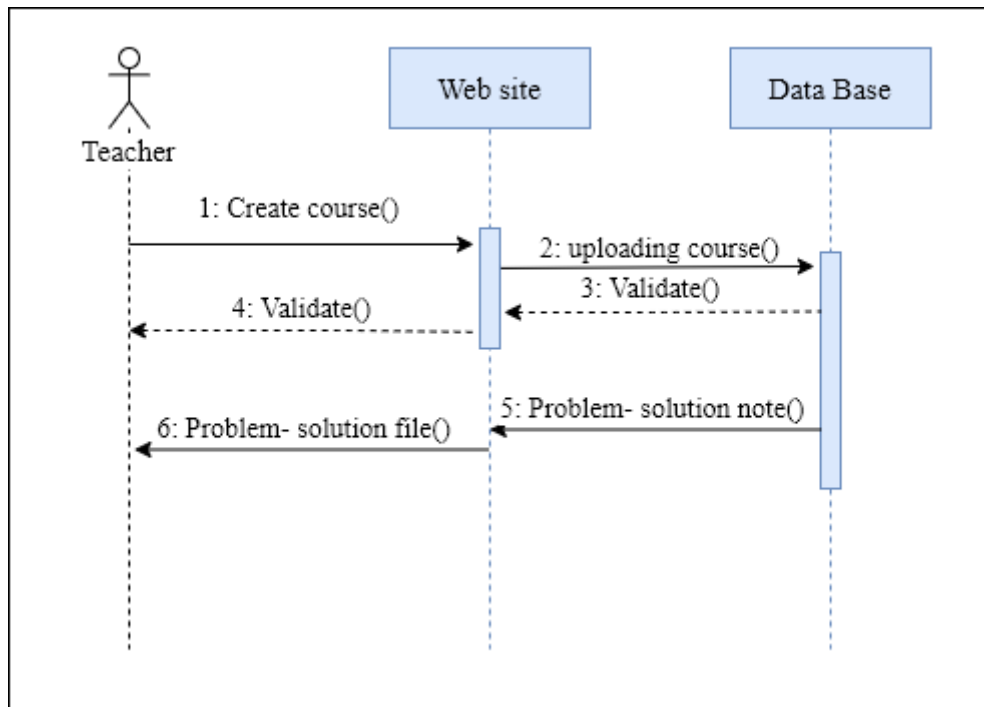


Figure 13: Sequence diagram of the teacher’s interactions with the platform.

9. System design

At the system design stage, the most widely used UML tool is class diagram. It is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.

The class diagram is used for general conceptual modeling of the structure of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed. The diagram above gives a global view of our project (Figure 14):

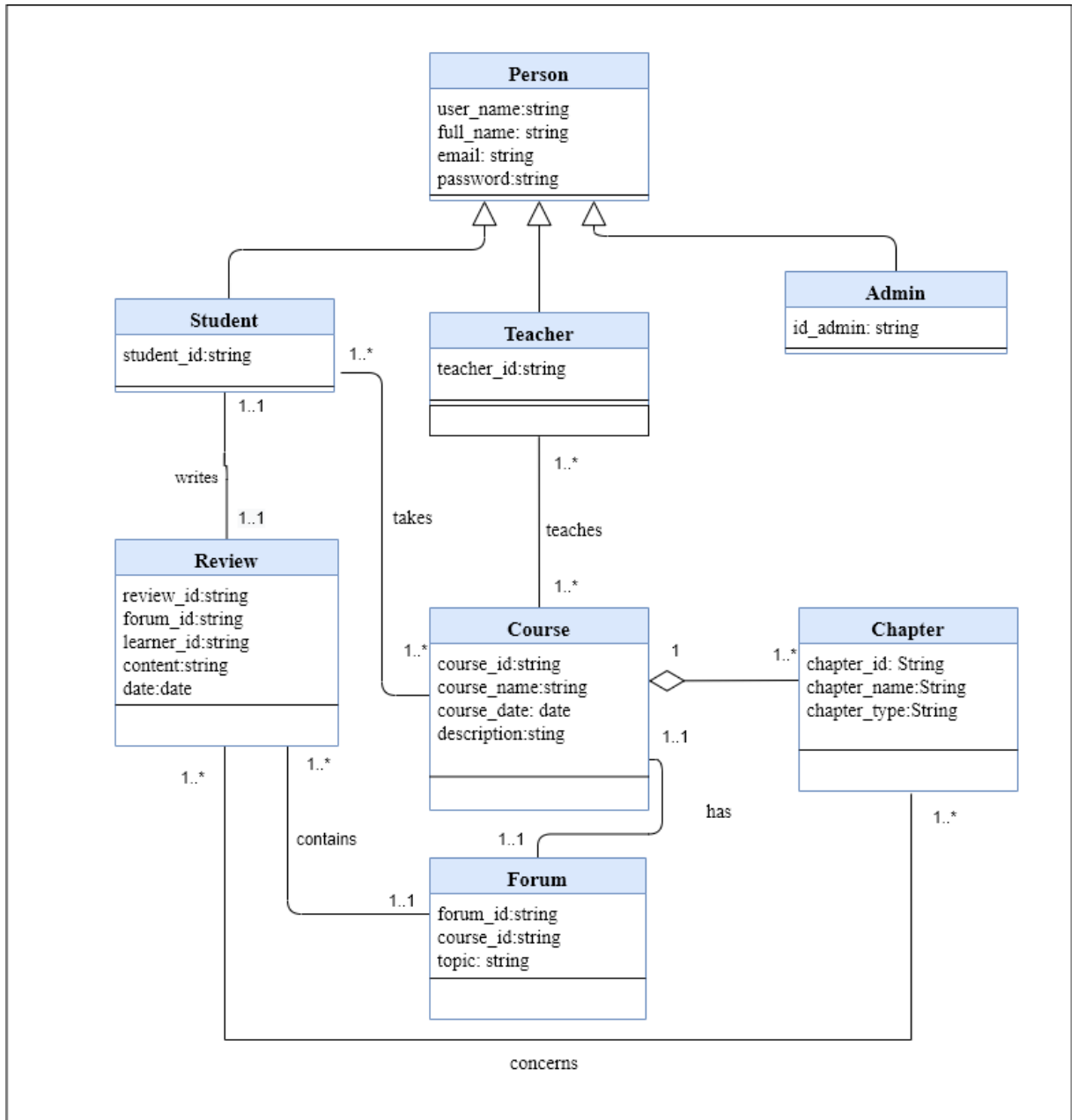


Figure 14: Class diagram of the platform.

Table 5 explains this diagram:

Table 6: Class diagram explanation.

Class	Code	Signification	Type
Person	user_name	User surname	String
	full_name	User full name	String
	email	User email	String
	password	User password	String
Student	id_student	Student identifier	String
Teacher	id_teacher	Teacher identifier	String

Admin	id_admin	Admin identifier	String
Course	course_id	Course identifier	String
	course_name	Couse name	String
	course_date	Course publication date	Date
	description	Description of the course and syllabus	String
Chapter	id_chapter	Chapter identifier	String
	chapter_name	Chapter name	String
	chapter_type	Type of the chapter (text, audio, video...)	String
Forum	forum_id	Forum identifier	String
	course_id	Course identifier	String
	topic	Topic of the forum	String
Review	review_id	Review identifier	String
	forum_id	Forum identifier	String
	id_student	Student identifier	String
	content	review content	String
	date	Review writing date	Date

Table 1: Class diagram explaining table

10. Conclusion

In this chapter we presented the development process as well as the conceptual vision and different functionalities of the platform.

The final chapter will be dedicated of the technical and practical side of the solution.

Chapter V:
Implementation of the solution

1. Introduction

Now that all the theoretical background related to our solution has been introduced and the modeling of the solution to the aim of enhancing student's experience in E-learning environments, we can talk more about the technical and practical aspects of the solution.

In this chapter, we will talk about implementation of our platform, the tools that we used, the programming languages and then we will present our platform.

2. Tools and languages used

In this section we will talk about the tools and programming languages we used to develop our platform.

- Python

Python¹² is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

Python is one of the most popular programming languages used by developers today¹³, and it is by far the most used language in the field of Artificial Intelligence. Most of the NLP and Machine Learning libraries are available in Python, and thus it is the best choice for our case.

¹² <https://www.python.org>

¹³ Number 1 in demand programming languages to learn in 2020 according to <https://towardsdatascience.com/>

- **Streamlit**

Streamlit¹⁴ is an open-source Python library that makes it easy to build custom web-apps for machine learning and data science. It is a simple and powerful app model that helps building rich UIs (User Interfaces) very fast.

- **Sublime Text**

Sublime¹⁵ text is a shareware cross-platform source code editor with a Python application programming interface (API). It supports many programming and markup languages, and functions. Moreover, it can be added by users with plugins, typically community-built and maintained under free-software licenses.

3. Architecture of the solution

Before presenting the platform, we will present briefly its most important components.

3.1. Datasets

Datasets are a collection of data. Dataset corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the dataset in question. And in our case we used two datasets.

3.1.1. Problem-solution Dataset

From ELearning disadvantage above-mentioned, we succeeded to detect the most common issues students face with these platforms. And we created our Problem-Solution dataset composed of:

1. Category problems
2. Problem keywords
3. Solutions

¹⁴ <https://www.streamlit.io>

¹⁵ <https://www.sublimetext.com/>

Category problem defines the category of the student problem. Problem keywords are the keywords of the problem category, since one category can be defined differently. And solution is the solution to the problem category detected.

And the table below (Table 7) gives more details about our dataset:

Table 7: Problem-solution dataset description.

Problem category	Problem keywords	Solution
Video problems	'video problems', 'clip', 'playback error', 'video unavailable', 'something went wrong'.	Please fix or re-do video!
Hardware and software problems / Technical issues	'slow computer', 'computer keeps restarting', 'driver issues', 'peripheral commands interpreted incorrectly', 'internet is slow', 'downloads take forever', 'computer freezes', 'attachments won't open', 'pc blue', 'pc blue screen of death', 'pop-up ads', 'corrupt files', 'unusual noises', 'graphical errors'.	Make the course multi-device.
Not enough exercises and examples	'examples issues', 'exercises issues', 'need more examples', 'need more exercises'	Add more exercises and/or examples.
Not detailed course	'not detailed enough', 'need a explication', 'need more details', 'misunderstanding',	Add more details and explanations.

	'reading materials missing'	
Audio problems	'audio issues', 'sound issues', 'voice', 'bad sound', 'hear', 'audio'.	Please fix or re-do audio.
Too many useless information	'boring', 'too long', 'too many useless information', 'unpractical'	Make the course more practical and entertaining.

3.1.2. Coursera reviews dataset

For the reviews part we used Coursera¹⁶ “100k-courseras-course-reviews-dataset”¹⁷ and took some reviews (around 30 reviews).

3.2. Database

A database is a data structure that stores organized information. Most databases contain multiple tables, which may each include several different fields. For the storage of our data, we used a database made of 3 different tables:

1. Userstable
2. Coursetable
3. Reviewstable

The table below (Table 7) gives more details about our database:

Table 8: Database description.

Table name	Code	Signification	Type
usertable	username	The username of the user is unique.	String

¹⁶ Coursera is an education platform that partners with universities and other organizations to offer online courses, specializations, and degrees in a variety of subjects,

¹⁷ 100K+ Scraped Course Reviews from the Coursera Website (<https://www.kaggle.com/septa97/100k-courseras-course-reviews-dataset>)

	password	The password of the user.	String
	fullname	The full name of the user.	String
	usertype	The type of the user (admin, teacher, student).	String
coursestable	courseid	The identifier of the course is unique.	Integer
	username	The username of user that created the course.	String
	title	The title of the course.	String
	description	The description of the course.	String
	postdate	The date of course publication.	Date
reviewstable	username	The username of the user who wrote the review.	String
	courseid	The identifier of the course the review is about.	String
	content	The content of the review	String

3.3. Data preprocessing and filtration

For data preprocessing and filtration processes of our platform, we used regular expression. Since our project is working with sentiment analysis we cannot use the usual NLP preprocessing function because we need the important words in the reviews to stay the same.

A regular expression (shortened as regex or regexp; also referred to as rational expression) is a sequence of characters that define a search pattern. Usually such patterns are used by string searching algorithms for "find" or "find and replace" operations on strings, or for input validation. It can be used in almost any coding language, and is very useful when trying to search for general string patterns. It is also used in web scraping, input validation, and simple parsing.

3.3.1. Regular Expression

A simple regular expression is one that is composed of just the literals. A literal is defined as a character that represents itself. Thus, the character "a" is expected to match itself in any text string but it may or may not match its upper case version "A".

In these examples, regular expressions are in a pair of / characters and the matching portions are underlined in the text:

A simple regular expression **/ring/** matches the following patterns:

The phone is ringing

Joe proposed to Mary with a ring.

And the regular expression **/day is/** matches:

Today is a Tuesday.

John said that his day is made.

The expression **/April 26, 1995/** matches:

John was born on April 26, 1995.

As the examples shows, a simple regular expression contains literals that can be alphabets, numbers, or special characters.

Now we will explain some basics about the use of meta-characters to add variable information into the regular expressions:

- The '.' character means match any single character.
- The '*' character means match zero or more of the thing before me. For example, 'a*' will match zero or more **a** characters. And '.'* means match zero or more of any characters, same as a single '*' would do in a wildcard expression.
- The '+' character means match one or more of the thing before me. 'a+' will match one or more **a** characters, but will not match if there are no **a** characters at all. '.'+' will match one or more of any characters.

And the table down below (Table 9) gives more details about the logic behind regular expressions:

Table 9: Regular expressions examples.

Regular expression	Match
[Pp]ython	"Python" or "python"
rub[ye]	"ruby" or "rube"
[aeiou]	Any one lowercase vowel
[0-9]	Any digit; same as [0123456789]
[a-z]	Any lowercase ASCII letter
[A-Z]	Any uppercase ASCII letter
[^aeiou]	Anything other than a lowercase vowel
[^0-9]	Anything other than a digit

In regular expressions, all characters match themselves. Except for control characters, + ? . * \$ ([] { } | \), but we can escape a control character by preceding it with a backslash “\”.

3.3.2. Data preprocessing

Since regular expressions are used in web scraping and input validation. We will use them in our preprocessing section to organize data.

In Python the regular expression package “Regex” can be imported using the command “*import re*”. This gives access to many different functions and string sequences that cleans data. One of the most important regular expression methods that we will use is **sub** that replace a string that matches a regular expression instead of perfect match.

Example:

Code:

```
import re
#review example
review="#Video unavailable?!$The Videos no longer exist123215"
reviews= re.sub('[#|*|$|:|\\|&|?|.|\'|...]', '', review)
print("review after removing special characters except exclamation point: ", review)
review= re.sub('[0-9]', '', review)
print("review after removing numbers: ", review)
```

Output:

#in the sentiment analysis section we will explain why we keep the exclamation points

The review after removing special characters except exclamation point:

Video unavailable!!The Videos no longer exist123215

The review after removing numbers:

Video unavailable!!The Videos no longer exist

3.3.3. Filtration

To detect eLearning problems from our dataset in the reviews, we have to use a filter that will be able to find the problems’ keywords in the reviews and giveback the problem category and suggest the solution.

And since similarity functions generally take in consideration only the closest concept. We used the regular expression methods in this step as well, because a review can talk about two or three eLearning problems.

Again we used “Regex’ and this time we used another regular expression methods that is *search()* that searches the string for a match, and returns a match object if there is a match.

Example:

Code:

```
import re
review = "Not bad! but we need more exercises"
x = re.search("exercise", review)
#check if there is a match
if x:
    print("We have a match!")
else:
    print("No match")
```

Output

We have a match!

And using this method with our problem-solution dataset keywords did give us a better result than using similarity function that we have experienced as a first trial.

Now that we talked about data preprocessing and filtration methods that we used in our solution, we will move to the sentiment analysis process.

3.4. Sentiment analysis

As for the sentiment analysis step, we used *textblob* functionalities. Since we are working with reviews and feedbacks that are simple sentences and all we want to know whether the review is positive or negative, *textblob* seems the most suitable option.

TextBlob is a Python library for processing textual data which is built on NLTK¹⁸ and Pattern. It provides a simple API for diving into common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation.

¹⁸ The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language.

In our case we used the sentiment property polarity that returns a float within the range [-1.0, 1.0], where 1 means positive statement and -1 means a negative statement.

The polarity of a sentence is calculated by using textblobs' Pattern Analyzer that assigns to each word of the sentence a polarity score. After this step, only the words with polarity different than "0.0" will be taken in consideration except for one punctuation mark ("!") . Exclamation point have "0.0" polarity but it does affect the polarity of the sentence depending on her polarity (if the sentence is negative he will make her more negative and if the sentence is positive he will make her more positive). After all these operations, the average polarity is calculated and it is returned as the polarity of the sentence.

The table and the example down below will provide more details about this function (Table 10):

Table 10: Words polarity and type according to textblob.

Word	Polarity score	Type
I	0.0	Neutral
Was	0.0	Neutral
Disappointed	-0.75	Negative
Because	0.0	Neutral
Of	0.0	Neutral
The	0.0	Neutral
Bad	-0.6999999999999998	Negative
Title	0.0	Neutral

Table 2: Words polarity and type according to textblob.

- **Example:**

Code:

```
from textblob import TextBlob
#use the sentence from the table as example
text = TextBlob("I was disappointed because of the bad title")
#show output
print("I was disappointed because of the bad title: ", text.sentiment.polarity)
#show another example with word Bad and the exclamation point
neg_word= TextBlob("Bad")
print("Bad : ", neg_word.sentiment.polarity)
neg_word= TextBlob("Bad!")
print("Bad! : ", neg_word.sentiment.polarity) #the polarity will change
```

Output:

```
#bad= -0.69  disappointed=-0.75

#I=0 ,was=0 ,because=0 ,of =0 ,the=0, title=0

# textblob(I was disappointed because of the bad title)= textblob(bad+
disappointed)

I was disappointed because of the bad title: -0.7249999999999999

Bad : -0.69999999999999998

Bad! : -0.87499999999999998
```

Now that we explained the important processes and the technical details about our platform, we can move to the presentation section.

4. Presentation of the platform

In the presentation section, we will show the interfaces of the platform as presented for different users through some screenshots, highlighting the main functions.

4.1. Home

Home or home page is the main web page of a website. It is the start page of our platform that gives a small presentation about what it does (Figure 15).

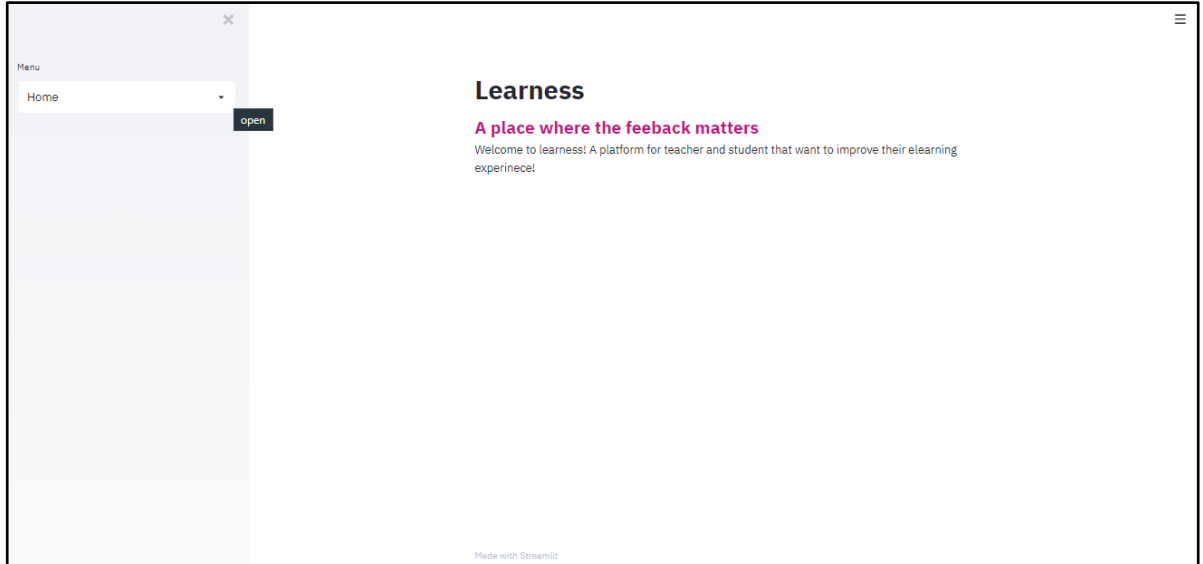


Figure 15: Learness home page.

4.2. Courses

After clicking on the menu and selecting “Courses” (Figure 16), the user will be able to see all the courses that are on the platform.

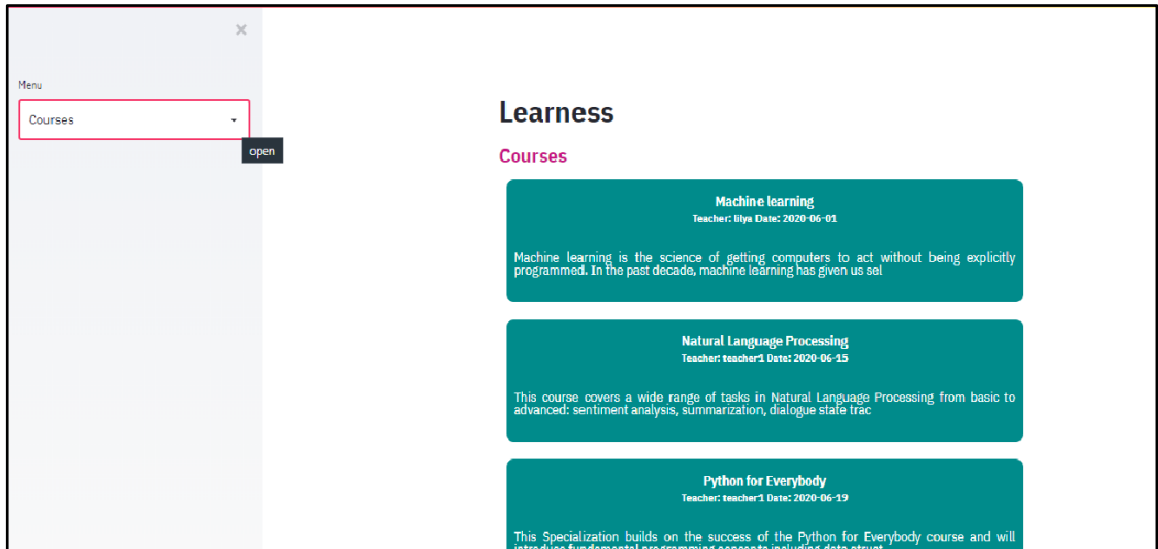


Figure 16 : Learness courses page.

4.3. Login

By clicking on the menu and selecting “Login”, the user will be able to login, in case he already has an account (Figure 17).

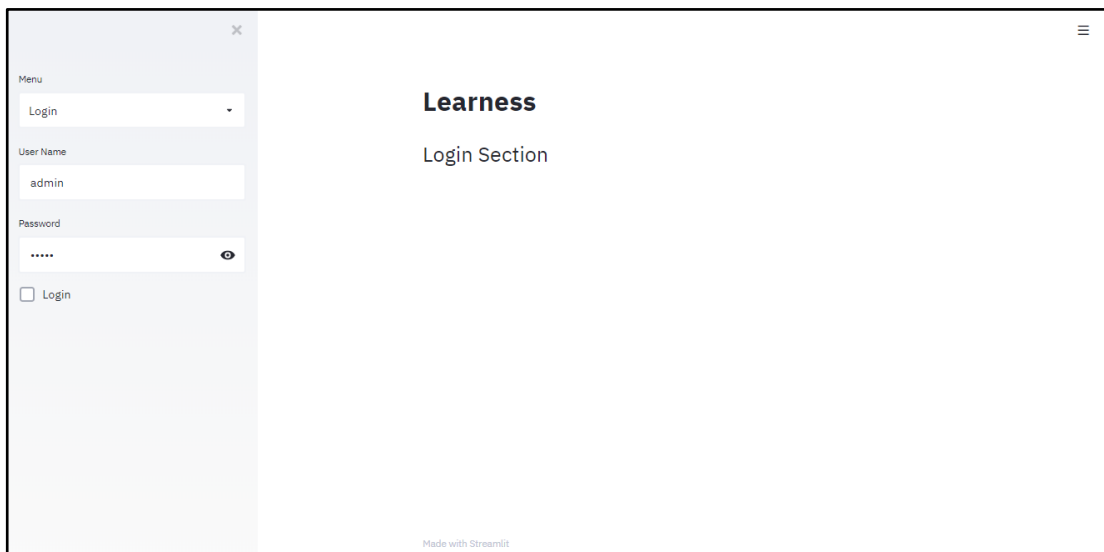


Figure 17: Learness login page.

In result, the user that just logged in will be directed to an interface that shows only the functionalities allowed accordingly to its type.

4.3.1. Administrator

Once logged in, the administrator can check user profiles by selecting the “Profiles” in the sidebar (Figure 18).

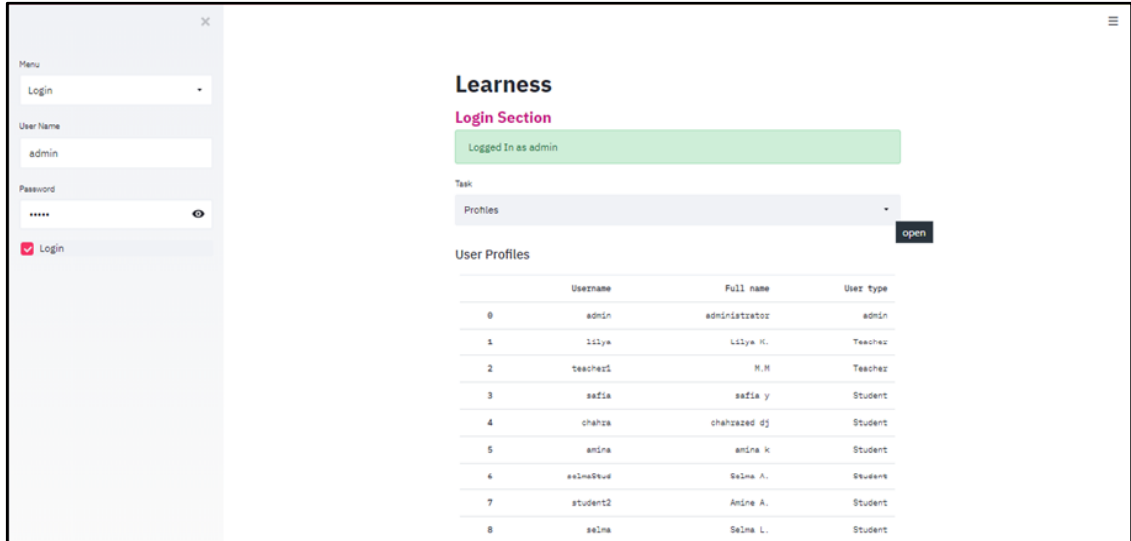


Figure 18: Administrator profiles page

He can also check the reviews table by selecting “Reviews” in the sidebar (Figure 19) and delete all the reviews by clicking on “Delete all” button.

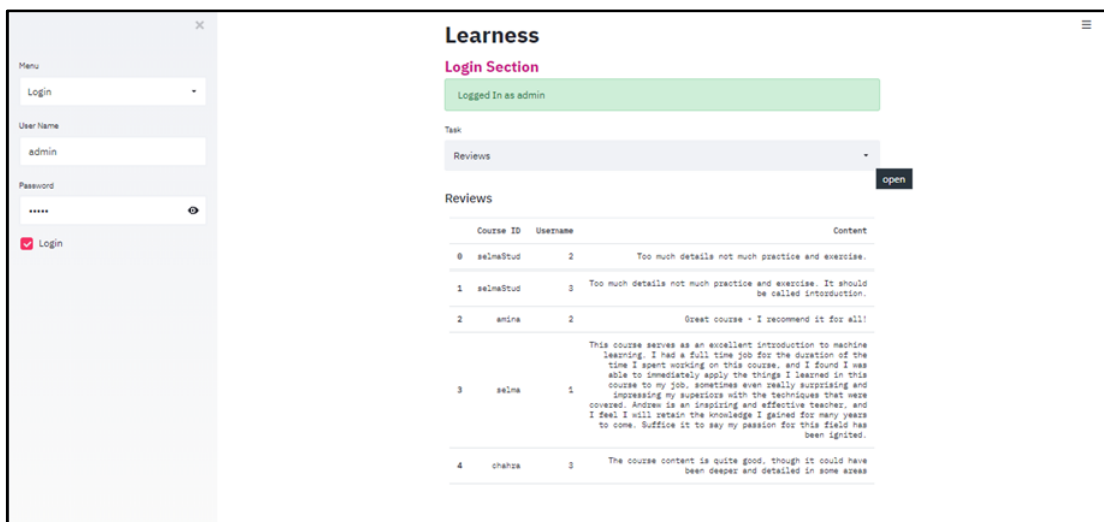


Figure 19: Administrator reviews page.

He can upload an excel file with reviews by selecting the “Reviews file” in the sidebar. After uploading the file, he can see the table and by clicking on the “Add reviews” the reviews will be added in the reviews table (Figure 20).

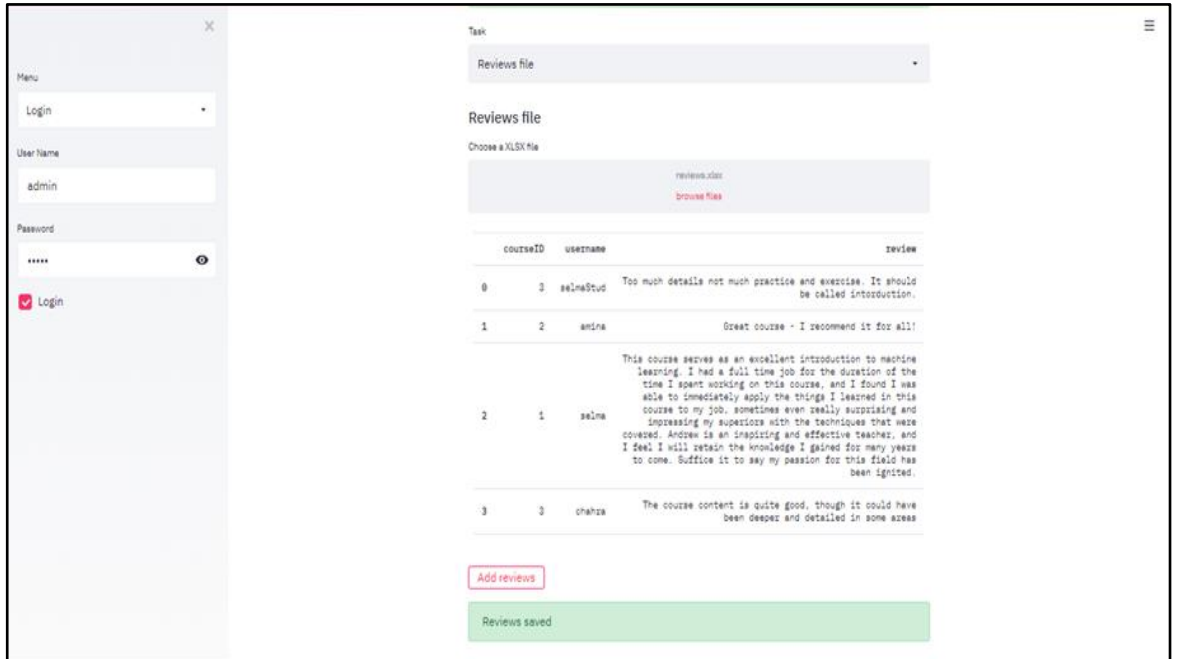


Figure 20: Administrator uploading reviews

And he can also check and see each of the steps of the solution process by selecting the “Steps” in the sidebar.

1. Preprocessing results (Figure 21).

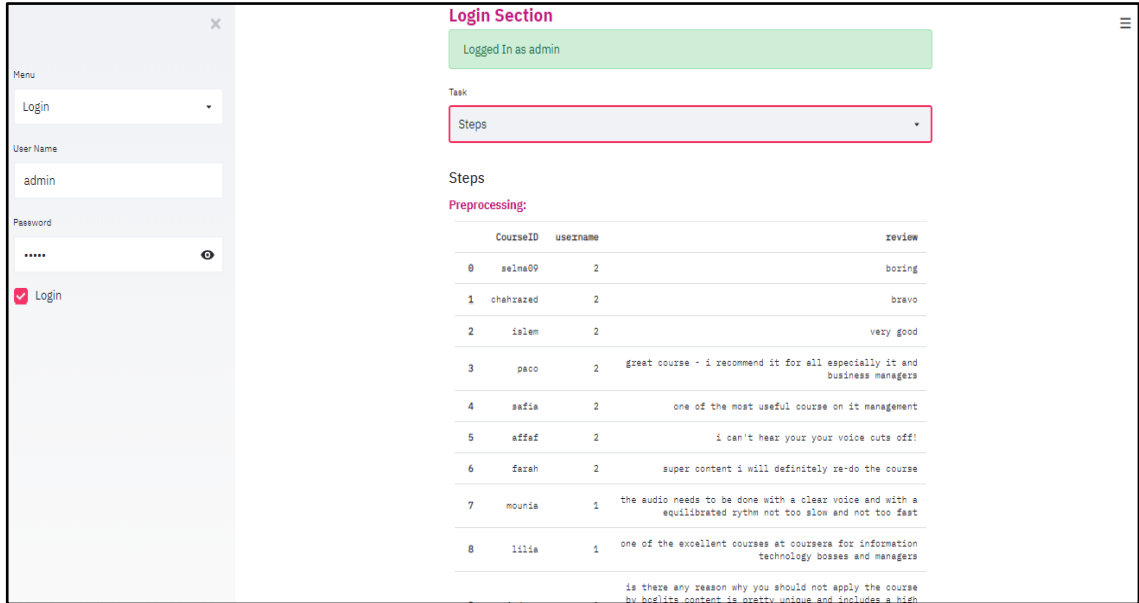


Figure 21: Admin checking preprocessing results.

2. Polarity detection results (Figure 22).

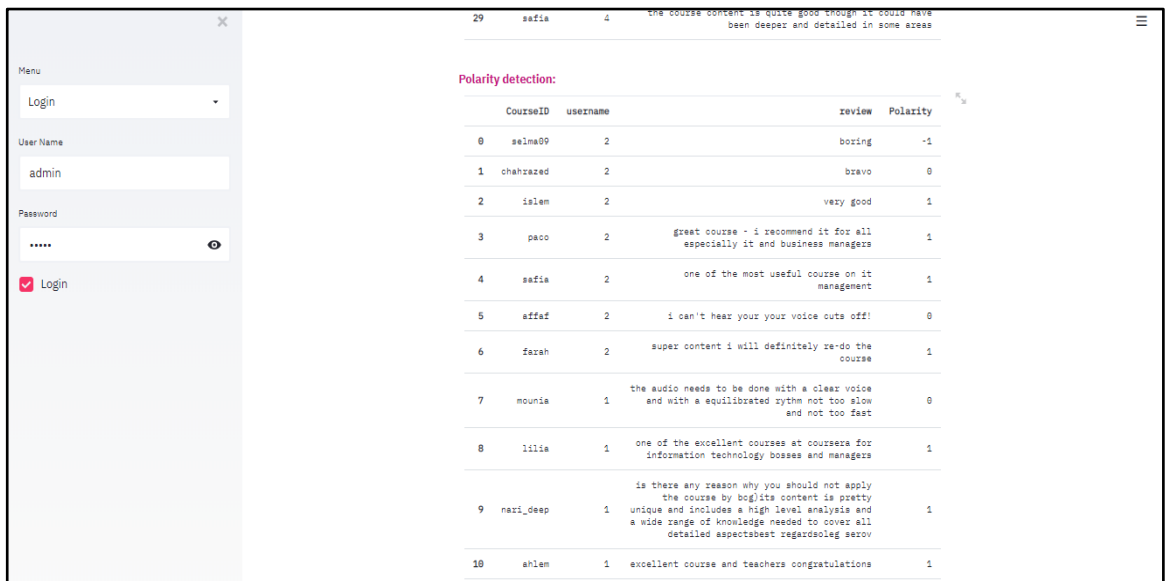


Figure 22: Admin checking polarity results.

3. After filtration results (Figure 23).

CourseID	username	review	Polarity	Problems	Solutions	
0	seina09	2	boring	-1	Too many useless information	Make the course more practical and entertaining.
1	chehrazed	2	bravo	0	default	there is no serious problems !
5	affaf	2	i can't hear your voice	0	Audio problems	Please fix or re-do audio.
7	mounia	1	the audio needs to be done with a clear voice and with a equilibrated rythm not too slow and not too fast	0	Audio problems	Please fix or re-do audio.
12	oualid	1	the course setting does (at least for me) not allow learn the content long term due to missing reading material	-1	Not detailed course	Add more details and explanations.
15	haman	1	too long!!!!!!!	-1	Too many useless information	Make the course more practical and entertaining.

Figure 23: Admin checking filtration results

4.3.2. Teacher

Once logged in, the teacher can create a new course by selecting the “Add course” in the sidebar (Figure 24). And since our platform is more focused about reviews and feedback, we did not expand all the functionalities regarding course creation. In this regard, the teacher can only give course title, description and the creation date.

Figure 24: Teacher adding new course.

He can also check the feedback, by selecting the “Feedback” in the sidebar (Our problem solution feedback). And after selecting his course title (since a teacher can create many courses), he can see the feedback (only if there were a reviews about it) (Figure 25).

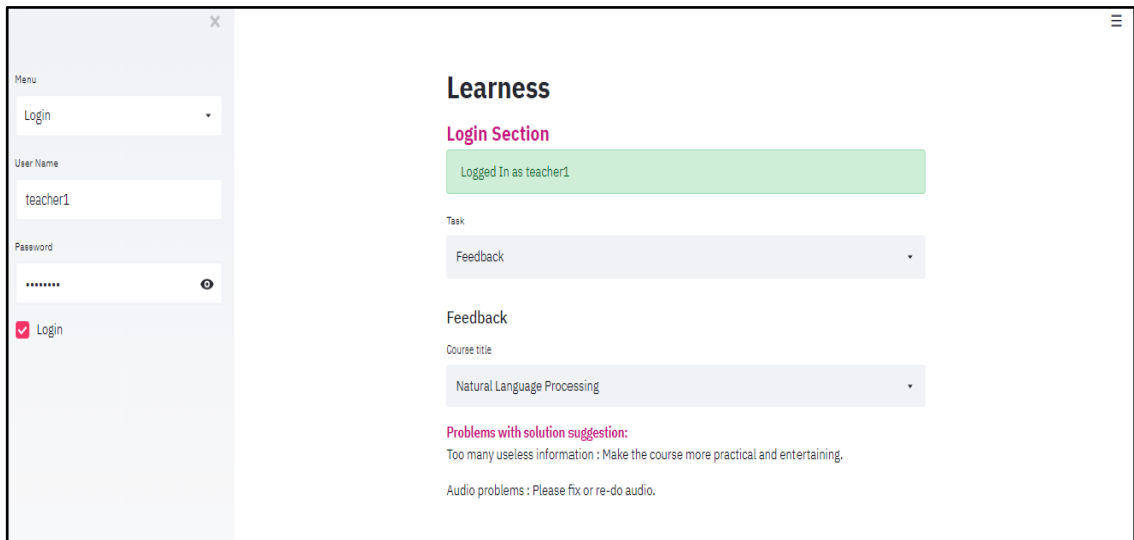


Figure 25: Teacher checking the feedback (unfinished).

4.3.3. Student

Once logged in, the student can check the course table by selecting the “Check course table” in the sidebar (Figure 26).

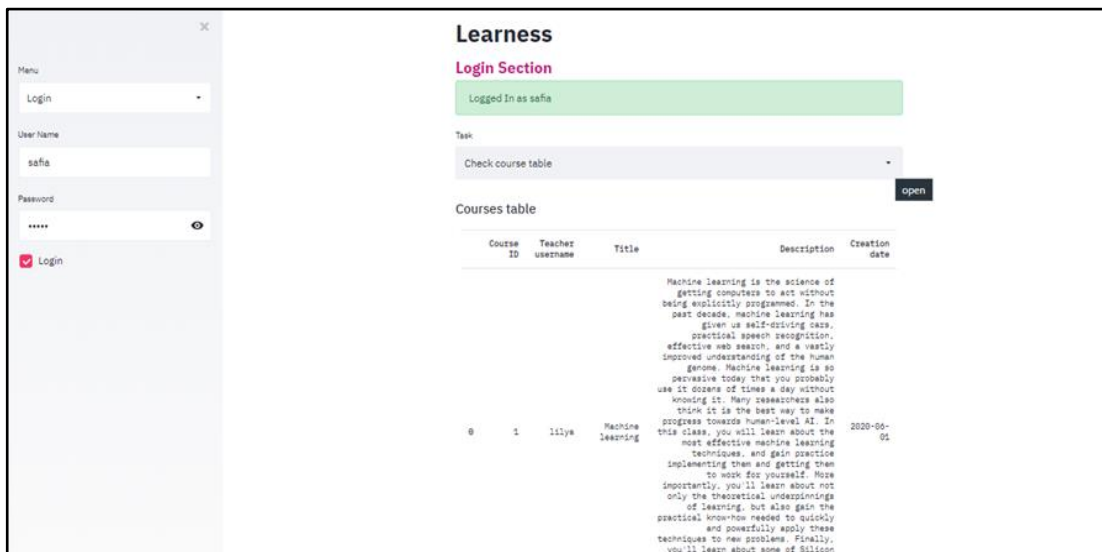


Figure 26: Student checking the course table.

He can also check the forum by selecting “Forum” in the sidebar, to see other students reviews (Figure 27).

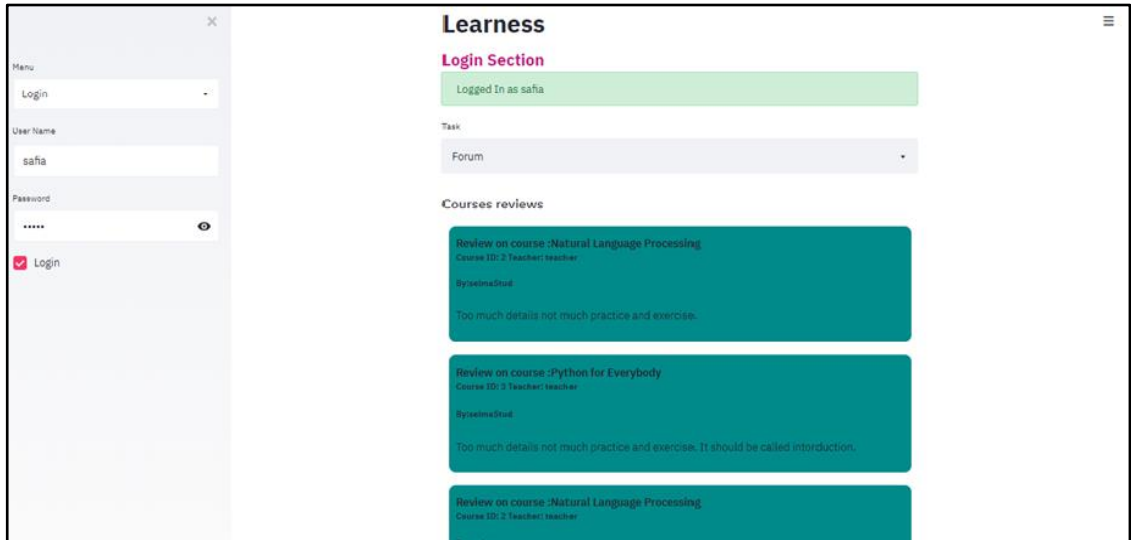


Figure 27: Student checking the forum.

And he can write a review by selecting the “Write review” in the sidebar. He has to select the course ID, check for the course title and teacher’s username, and then write the comment and click on “Add review” to add the review (Figure 28).

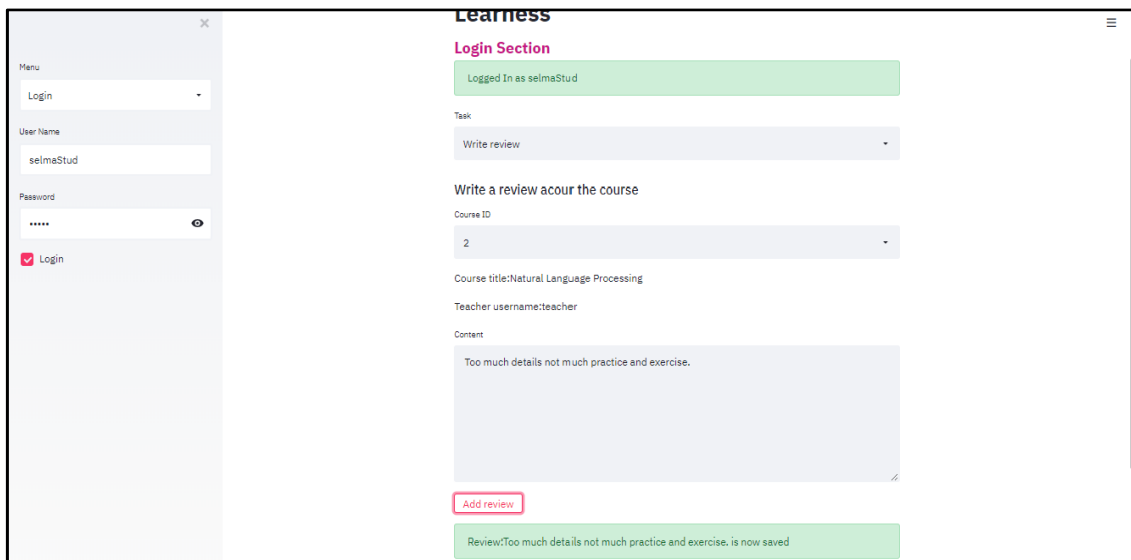


Figure 28: Student writing a review.

5. Signup

To sign up and create a new account the user has to select “SignUp” in the sidebar (Figure 29). And choose a unique username, a password, write his full name and choose his user type (Teacher or Student).

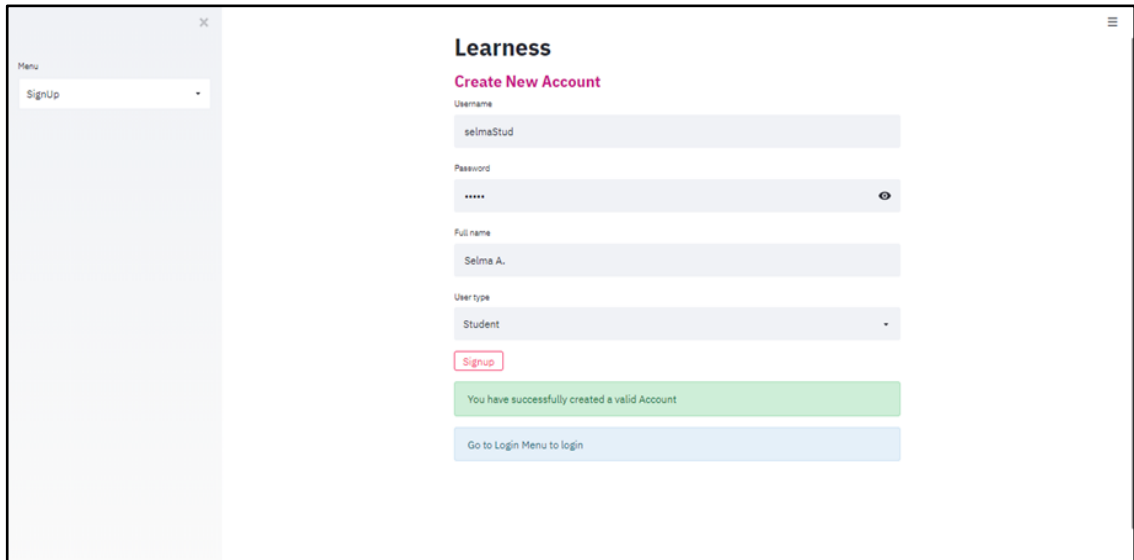


Figure 29: SignUp page.

6. Test and evaluation

In this section we will explain the actions performed as for the testing and evaluation part. Test is the process that ensures that the platform meets requirements and it is working properly. Evaluation on the other hand is the process of judging how well the originally intended objectives have been achieved.

6.1. Test part

In the beginning, we choose randomly 30 reviews from coursera dataset. Then, we create users (teachers and student) and courses that the reviews will refer to.

After this, we create an excel file with course ID, username and the reviews and, as an admin, we upload this reviews file. Then the program will do the preprocessing, sentiment analysis and filtration processes, and returns back the feedbacks (problems and solution suggestion accordingly to the courses' reviews).

Finally, we checked the feedback as teacher of the course (Figure 30).

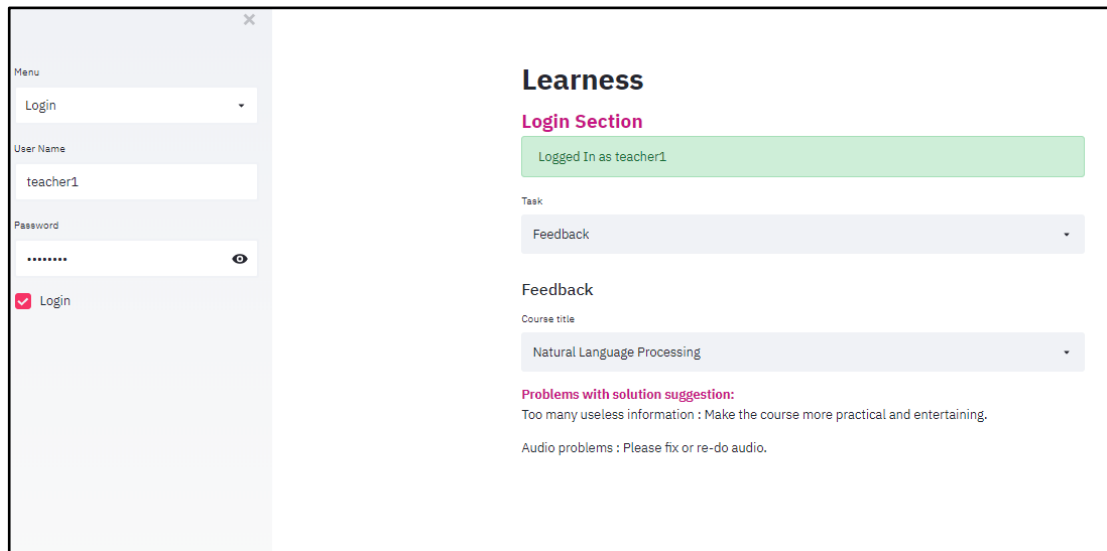


Figure 30: Checking the feedback as teacher

6.1.1. Evaluation part

In this part, we present the performance of our solution. For that first, we have to check the reviews and detect the polarity and the problems manually then compare them with the obtained results.

We observed that the system performed well in 78.66% cases and 78.33% case for negative neutral reviews detections and problems detection respectively.

More accurately, we took five random courses and wanted to further the performance tests, by computing the well-known performance measures “Precision” and “Recall”.

As a reminder, the calculation of these measures is based on the comparison between the detections produced automatically by our solution which we will call “S” and a set of reference detection produced by a human which we will denote “H”.

- Correct detections found by the system are called "the true positive (TP)" and are calculated as follows:

$$TP = S \cap H$$

- Incorrect detections found by the system are called "the false positive (FP)" and are calculated as follows:

$$FP = S - S \cap H$$

- Correct detections omitted by the system are called "the false negative (FN)" and are calculated as follows:

$$FN = H - S \cap H$$

In this regard:

- Precision is a measure of accuracy, it varies between [0,1] and is calculated as follows:

$$Precision = \frac{|TP|}{|TP + FP|}$$

- The recall is a measure of perfection, it varies between [0,1], it is calculated as follows:

$$Recall = \frac{|TP|}{|TP + FN|}$$

The values obtained by calculating successful detections are not comparable through precision and recall, in fact, the recall can take on significant values at the expense of precision, returning all possible detections.

At the same time, precision can take on important values at the expense of recall, while returning that the correct detections even if they are few.

It is for these reasons that it is preferable to take into account the two measurements simultaneously via a measurement which combines recall and precision such as: the F-score which is calculated as follows:

$$F - score = \frac{2 * (Recall * Precision)}{(Precision + Recall)}$$

F-score is a global measure of the quality of successful detections, it varies between [0.1], and this measure assigns the same importance to precision and recall.

The table down below, summarizes the obtained results.

First, Table 11 presents an overview of the detections obtained by the system and those that we have detected.

Table 11: Detections performed by the system VS Detections performed by human.

Course ID	Negative-neutral reviews detected / negative neutral comment		Problems detected / problems	
	Solution	Human	Solution	Human
1	3	6	3	4
2	3	2	2	2
3	4	4	2	2
4	3	5	2	3
5	1	2	1	2

And the performance measurements obtained for “Negative-neutral reviews” detections and “Problems” detections are summarized in tables 12 and 13, respectively.

Table 12: Negative-neutral review detection evaluation.

Course ID	Total Reviews / Course	TP	FP	FN	Precision	Recall	F-score
1	10	3	3	0,67	0,50	1,00	0,67
2	7	2	0	0,80	1,00	0,67	0,80
3	4	4	0	1,00	1,00	1,00	1,00

4	3	3	2	0,75	0,60	1,00	0,75
5	6	1	1	0,75	0,50	1,00	0,67
Medium Precision, Recall, and F-score					0,72	0,93	0,81

Table 13: Problems detection evaluation.

Course ID	Total Reviews / Course	TP	FP	FN	Precision	Recall	F-score
1	10	3	1	0	0,75	1,00	0,86
2	7	2	0	0	1,00	1,00	1,00
3	4	2	0	0	1,00	1,00	1,00
4	3	2	1	0	0,67	1,00	0,80
5	6	1	1	0	0,50	1,00	0,67
Medium Precision, Recall, and F-score					0,78	1,00	0,88

And to see this more clearly, we will take 3 students reviews as an example:

1. **Review 1:** “Very Gooooood !! “
2. **Review 2:** “I can't hear you, your voice cuts off!!!!!!”
3. **Review 3:** “sorry but we can't understand , we need more examples please !”

After the preprocessing it will be:

1. **Review 1:** “very good”
2. **Review 2:** “I can t hear you your voice cuts of !”
3. **Review 3:** “sorry but we can t understand , we need more examples please !”

After the sentiment analysis step it will transform into a set of preprocessed review and it polarity(1when it is positive, 0 when it is neutral and -1 when it is negative):

1. **Review 1:** (“very good”, 1)
2. **Review 2:** (“I can t hear you your voice cuts of !”, -1)

3. **Review 3:** (“sorry but we can t understand , we need more examples please !” , -1)

And after the filtration step (the ELearning problem detection) which contains only the negative or neutral reviews (Since Review 1 is positive it will not be used) we will have another set as result(preprocessed reviews, their polarities and the ELearning problem detected):

1. **Review 2:** (“I can t hear you your voice cuts of !”, -1, “Audio problems”).
2. **Review 3:** (“sorry but we can t understand , we need more examples please !” , -1,“Not enough exercises and examples”).

And the feedback (Problem – Solution) the teacher will receive will be:

1. **From Review 2:** Audio problems - Please fix or re-do video.
2. **From Review 3:** Not enough exercises and examples - Add more exercises and/or examples.

Through the two tables and the examples , we can notice that the results obtained were rather encouraging at least for this set of tests. Therefore, it would be interesting to extend the tests to other documents to have a better appreciation of the results.

7. Conclusion

In this chapter, we have showed the implementation side of our solution focusing on the component’s functionalities, the platform presentation, and finally the testing and evaluation of the results. The obtained results of our tests are very encouraging and motivate us to explore more the use of sentiment analysis in eLearning field

General conclusion

1. Summary

Applying a sentiment analysis to examine the students' reviews in eLearning platform turns out to be an important endeavor. However, the current accuracy is promising for effective analysis of students' reviews sentiments. Such analysis can help provide a better understanding of users' opinions and issues regarding the e-learning system for the sake of its improvement.

In this thesis we demonstrated the promising attempt of using sentiment analysis in eLearning problem detection, which helps to improve the teaching-learning process. In particular we focused on proposing an automatized feedback method so the confidentiality of the students' communications will be ensured.

By applying three selected features (sentiment analysis, data processing and filtration) on eLearning reviews. To validate our solution, we analyzed students' reviews and comments obtained from the Coursera eLearning platform for different courses and compared the eLearning problems students had about each course. Then, sent the problem detected with potential solution suggestion to the teacher. This solution has demonstrated the promising attempt of incorporating sentiment classification and problems detection into e-learning.

2. Outlooks

For future work, one main direction is to investigate whether it is possible to create the problem-solution dataset from the reviews, try another sentiment analysis methods and other forms of feedback (not only textual). Also suggest a modified subjectivity corpus to be used in the academic domain to achieve better results.

Moreover, we think it will be interesting to test this methodology for other languages than English and why not propose a Moodle (or other platform) plugin that can be integrated and used directly in existing Elearning platforms.

References

- [1] Z. a. b. a. M. a. A. A. Kechaou, *Improving e-learning with sentiment analysis of users' opinions*, 2011.
- [2] ElearningNS, 2020. [Online].
- [3] J. a. G. M. Williams, *The evolution of e-learning*, 2005.
- [4] A. S. a. X. Chen, *Online Education and Its Effective Practice*., vol. 15, 2016.
- [5] K. & E. Neuböck, *publikationen rund um das pionierprojekt*.
- [6] M. Mohammad Khalil, *Learning Analytics in Massive Open*, 2017.
- [7] L. E. H. H. F. T. a. M. Evelyn Kigozi Kahiigil, *Exploring the e-Learning State of Art*, vol. 6, 2014.
- [8] E. L. D. D. S. T. M. L. B. K. S. W. & C. J. D. Meyen, *An academy: Report on linking teacher education to advances in research*, 2002.
- [9] M. A. KARA Mostefa, *Analyse des sentiments dans les réseaux*, 2018 – 2019.
- [10] Classter, 2020. [Online].
]
- [11] N. Guragain, *E-Learning Benefits and Applications*, 2016.
]
- [12] F. BELBACHIR, *Le e-learning comme méthode d'apprentissage*, 2016.

]

[13 F. a. A. O. o. t. U. N. (FAO), *learning and facilitation tools*, 2011.

]

[14 B. L. Quality, *Criteria for an eLearning Platform*, 2013.

]

[15 T. & C. S. & X. G. & Z. J. Luo, *Sentiment Analysis*, 10.1007/978-1-4614-7202-5_4., 2013.

[16 M. W. & al., *Sentiment analysis algorithms and applications: A survey*, Ain Shams Eng J , 2014.

[17 D. Buzic, *Sentiment Analysis of Text Documents*, 2019.

]

[18 A. & F. F. & G. P. & G. T. D'Andrea, *Approaches, Tools and Applications for Sentiment Analysis Implementation. International Journal of Computer Applications*, 2015: 125. 26-33 10.5120/ijca2015905866..

[19 S. & M. T. & M. E. & A. I. & P. S. Jayasanka, *Sentiment Analysis for Social Media*, 2013.

[20 E. A. L. & al., *Evaluation of ELearning activity effectiveness in higher education through sentiment analysis by using Naïve Bayes classifier*, 2018.

[21 S. & K. P. Rani, *A Sentiment Analysis System to Improve Teaching and Learning*, 50. 36-43. 10.1109/MC.2017.133., 2017.

[22 M. & Y. D. & R. C. Wen, *Sentiment analysis in MOOC discussion forums: What does it tell us?*, Proceedings of Educational Data Mining. 130-137., 2014.

- [23 B. K. K. Singh, *Testing UML Designs using Class, Sequence and Activity Diagrams*, IJRST –International Journal for Innovative Research in Science & Technology| Volume 2 | Issue 03 |, 2015.
- [24 M. Danubianu, *STEP BY STEP DATA PREPROCESSING FOR DATA MINING. A CASE STUDY.*, 2015.
- [25 W. Bhaya, *Review of Data Preprocessing Techniques in Data Mining*, . Journal of Engineering and Applied Sciences. 12. 4102-4107. 10.3923/jeasci.2017.4102.4107. , 2017.
- [26 J. H. Gautami Tripathi, *FEATURE SELECTION AND CLASSIFICATION APPROACH FOR SENTIMENT ANALYSIS*, vol. 2, june 2015.
- [27 T. Slimani, *Description and Evaluation of Semantic Similarity Measures Approaches*, International Journal of Computer Applications. Vol 80. 25-33. 10.5120/13897-1851, 2013.
- [28 O. & O. K. Stephen, *Software Development Methodologies: Agile Model Vs V-Model.*, International Journal of Engineering and Technical Research. 2. 108-113. , 2014.
- [29 M. H. BENMOUSSA, *Développement d'une application de gestion d'un parc informatique*, 2014.
- [30 A. & K. A. & M. S. & S. S. & M. V. Guin, *Sentiment Analysis for Text*,] 10.13140/RG.2.1.4434.8400. , 2015.
- [31 H. & ., R. & L. M. Rana, *E-learning: Issues and Challenges*, . International Journal of Computer Applications. 97. 20-24. 10.5120/17004-7154. , 2014.
- [32 L. E. H. H. F. T. a. M. Evelyn Kigozi Kahiigi, *Exploring the e-Learning State of*

] *Art*, vol. 6 , 2008.