People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
ECHAHID CHEIKH LARBI TEBESSI UNIVERSITY - TEBESSA
Faculty of Exact Sciences and Natural Sciences
Department of Computer Science
Specialty:Systems and multimedia

Graduation Project for obtaining Master Degree

Machine learning-Based Approach to Annotate Social Media Comments for Product Recommendation

Created by:
Zouaoui Wafa

The Jury Members:

• Supervisor	Brek Assia	Dr	University of Tebessa
• Co-supervisor	Laifa Hassiba	Dr	University of Tebessa
• President	Mahmoudi Rachid	Dr	University of Tebessa
• Examiner	Chergui Othaila	Dr	University of Tebessa

Date of defense: 09/06/2025

Scholar Year: 2024 / 2025

Acknowledgments

First, I thank God for granting me success and the will and patience to complete my master's studies and accomplish this work.

I extend my sincere thanks to my thesis supervisor, **Dr. Brek Assia**, and co-supervisor **Dr. Laifa Hassiba**, for their valuable guidance, advice, and continuous support throughout this project.

I also extend my sincere thanks to all the professors who shared their expertise and supported me throughout my academic career.

I also extend my sincere thanks to my colleagues and friends for their moral support and constant encouragement.

Finally, I extend my sincere gratitude to my family, especially my parents, for their endless love, sacrifices, and continuous support throughout my journey.

Dédicaces

الحمد لله حبا وامتناناً على البدء والختام

بكل حب أهدي ثمرة نجاحي وتخرجي وحصاد ما زرعته سنيناً طويلة في سبيل العلم لـ :

الى من لا يفصل اسمه على اسمي ، ذلك الرجل العظيم، إلى من علمني أن الدنيا كفاح وسلاحها العلم والمعرفة، إلى من غرس في روحي مكارم الأخلاق، داعمي الأول وسندي، إلى من سعى لأجل راحتي ونجاحي إلى أعظم رجل في الكون إلى فخري واعتزازي

أبي الغالي

إلى أنيسة العمر و حبيبة الروح وأعظم نعم الله علي، إلى التي ضمت اسمي بدعواتها في ليلها ونهارها وأضاءت بالحب دربي، وأنارت باللطف والود طريقي، وكانت في سحاباً ماطرا بالحب والعطاء إلى من كانت ملجأي ويدي اليمين في هذه المرحلة.

أمي الغالية

إلى السند والكتف الثابت الذين إذا مالت الدنيا لا يميلون، ملهمي نجاحي، إلى من شددت عضدي بهم فكانوا لي ينابيع أرتوي منها، إلى خيرة أيامي وصفوتها، إلى قرة عيني، القلب الدافئ والابتسامة الصادقة، أول الداعمين

خواتي "ولاء و امال " و اخواني " اشرف واسلام "

إلى توأم الروح وبهجة الأيام، إلى من كانوا السند أوقات الضعف، من شاركوني التعب والفرح طيلة المشوار اخواتي التي لم تلدهم امي إلى توأم الروح وبهجة الأيام، إلى من كانوا السند أوقات الضعف، من شاركوني التعب والفوح طيلة المشوار اخواتي التي لم تلدهم امي

"اسراء اميرة اية صفاء رحمة خلود نورهان و وفاء رحاب عبير "

إلى " نفسي "

بعد سنوات من الجهد والسهر والتحديات أقف اليوم بفخر أمام إنجاز لطالما حلمت به
لقد آمنت بقدرتي رغم التعب ، رغم ثقل الطريق، رغم كل الصعاب التي واجهتها

هذا التخرج ليس نهاية، بل بداية لطريق جديد..

Abstract

Sentiment analysis is a core task in Natural Language Processing (NLP), widely applied in sectors such as digital marketing, politics, and customer feedback systems. However, its application to Arabic dialects—particularly the Algerian dialect—remains challenging due to linguistic diversity, lack of standardized orthography, and the prevalence of non-standard scripts like Arabizi. Existing works are often limited to Modern Standard Arabic or fail to effectively capture dialect-specific nuances.

To address these limitations, this work proposes a late fusion architecture that combines contextual embeddings from two complementary pre-trained models: DziriBERT, specialized for Algerian dialect, and MARBERT, trained on a broad set of Arabic dialects. The fused representations are then passed to a BiLSTM classifier, selected for its strong capacity to capture sequential context.

Experiments conducted on two real-world datasets—Djezzy and BrandDZ—demonstrated the effectiveness of this approach. The BiLSTM-based late fusion model achieved accuracies of 0.83% , respectively, These results confirm that integrating dialectal and broader Arabic representations, along with appropriate sequence modeling, leads to more robust and reliable sentiment classification for under-resourced dialects such as Algerian Arabic.

Keywords: NLP, Sentiment analysis, Algerian dialect, Arabic dialects, Late fusion, Arabizi

Résumé

L'analyse des sentiments est une tâche fondamentale du traitement automatique du langage naturel (TALN), largement utilisée dans des domaines tels que le marketing digital, la politique et les systèmes de retour client. Toutefois, son application aux dialectes arabes—en particulier le dialecte algérien—reste difficile en raison de la diversité linguistique, de l'absence de normes orthographiques, et de l'utilisation fréquente de scripts non standard comme l'Arabizi. Les travaux existants sont souvent limités à l'arabe standard moderne ou ne parviennent pas à saisir efficacement les spécificités des dialectes.

Pour répondre à ces limites, ce travail propose une architecture de fusion tardive combinant les représentations contextuelles de deux modèles préentraînés complémentaires : DziriBERT, spécialisé dans le dialecte algérien, et MARBERT, entraîné sur un ensemble large de dialectes arabes. Les représentations fusionnées sont ensuite transmises à un classifieur BiLSTM, choisi pour sa capacité à capturer le contexte séquentiel.

Les expériences menées sur deux jeux de données réels—Djezzy et BrandDZ—ont démontré l'efficacité de cette approche. Le modèle basé sur la fusion tardive avec BiLSTM a atteint des précisions de 0,83% , Ces résultats confirment que l'intégration de représentations dialectales et arabes diversifiées, associée à un modèle séquentiel adapté, permet une classification des sentiments plus robuste et fiable pour les dialectes sous-ressourcés comme l'arabe algérien.

Mots-clés: NLP, Analyse de sentiment, Dialect Algérien,dialectes arabes, , Fusion tardive, Arabizi

الملخص

تحليل المشاعر هو مهمة أساسية في مجال معالجة اللغة الطبيعية، ويُستخدم على نطاق واسع في مجالات مثل التسويق الرقمي، والسياسة، وأنظمة تقييم رضا العملاء. ومع ذلك، فإن تطبيقه على اللهجات العربية – وخاصة اللهجة الجزائرية – لا يزال يمثل تحديًا بسبب التنوع اللغوي، وغياب معايير إملائية موحدة، والانتشار الواسع لنصوص مكتوبة بالفرانكو-عربي (Arabizi). وغالبًا ما تقتصر الأعمال السابقة على اللغة العربية الفصحى أو تفشل في التقاط الخصائص الدقيقة للهجات المحلية.

لمعالجة هذه التحديات، نقترح في هذا العمل نهج دمج متأخر يجمع بين التمثيلات السياقية لنموذجين لغويين مدربين مسبقًا ومتكاملين: DziriBERT المتخصص في اللهجة الجزائرية، وMARBERT الذي يغطي مجموعة واسعة من اللهجات العربية. يتم تمرير التمثيلات المدمجة إلى مصنف BiLSTM، الذي تم اختياره لقدراته العالية على فهم السياق التسلسلي في النصوص.

وقد أُجريت التجارب على مجموعتي بيانات حقيقيتين: Djezzy وBrandDZ، وأظهرت النتائج فعالية النموذج المقترح، حيث حقق دقة بلغت 0.83%، تؤكد هذه النتائج أن دمج تمثيلات لغوية متنوعة مع بنية تصنيف ملائمة يُعد خطوة فعالة نحو تحليل مشاعر أكثر دقة واستقرارًا للهجات الضعيفة الموارد مثل اللهجة الجزائرية.

الكلمات المفتاحية: معالجة اللغة الطبيعية، تحليل المشاعر، اللهجة الجزائرية، اللهجات العربية، الاندماج المتأخر، العربية

Table of contents

Acknowledgments **i**

Dédicaces **ii**

Abstract **iii**

Résumé **iv**

Abbreviations **x**

General introduction **1**

1 Background **3**

1.1 Introduction 4

1.2 Natural Language processing 4

1.2.1 Definition 4

1.2.2 Historique du NLP 5

1.2.3 NLP branches 7

1.2.4 NLP in action 7

1.2.5 NLP Applications 11

1.2.6 NLP advancements 12

1.3 Arabic language 13

1.3.1 Classic Arabic (CA) 13

1.3.2 Modern Standard Arabic (MSA) 13

1.3.3 Dialectal Arabic 13

1.4 Algerian dialect (AD) 14

1.4.1 Characteristics of Algerian Dialect 14

1.4.2 Difficulties of Arabic language processing 15

1.5 Conclusion 15

2 Related work **16**

2.1 Introduction 17

2.2 Sentiment Analysis 17

2.2.1 Definition of Sentiment Analysis 17

2.2.2 Types of Sentiment Analysis 17
2.2.3 Sentiment Analysis Levels 19
2.3 Sentiment analysis methods 19
2.3.1 Lexicon-based approach 20
2.3.2 Traditional machin learning classifiers 21
2.3.3 Deep learning-based classifiers 24
2.3.4 Tansformer-based models 26
2.4 Sentiment Analysis Domain of Application 27
2.5 The Challenges of Sentiment Analysis : 28
2.6 E-Commerce and Recommendation Systems 29
2.6.1 Overview and Definition recommendation Systems 29
2.6.2 Role of Sentiment Analysis in Enhancing Recommendation Systems 29
2.7 Comparison study. 30
2.7.1 Case of Arabic Algerian dialect (Arabic Script) 30
2.7.2 Arabic Algerian Dialect (Latin or Arabic script) 33
2.7.3 Case of Modern Standard Arabic 37
2.8 Conclusion 40

3 Methodology and Results **41**
3.1 Introduction 42
3.2 Contribution 42
3.2.1 Data overview 44
3.2.2 Data Preprocessing and preparing : 46
3.2.3 Propostion 48
3.2.4 Evaluation metrics 50
3.2.5 Results Analysis 51
3.3 Ablation study 54
3.4 Conclusion 57

General Conclusion **58**

List of Figures

1.1 Position of NLP in AI [1]. 5
1.2 Historique du NLP [2]. 6
1.3 NLP in action [4]. 7

2.1 Sentiment analysis grades [11]. 18
2.2 Emotion classes in sentiment analysis[11]. 18
2.3 Sentiment analysis methods[14] 20
2.4 Some Applications of Sentiment Analysis [36]. 27

3.1 Global proposed architecture. 43
3.2 Pre-treatment process 47

List of Tables

2.1 Case of Arabic Algerian dialect (Arabic Script) 32
2.2 Case of Arabic Algerian Dialect (Latin or Arabic script) 35
2.3 Case of Modern Standard Arabic 39

3.1 Datasets collection metadata 44
3.2 Dataset "Djezzy " 45
3.3 Dataset "BrandtDz " 45
3.4 Structure of the "Djezzyy" dataset 46
3.5 Structure of the "BrandtDz" dataset 46
3.6 Results of dataset "djezzy " with Translation 52
3.7 Results of dataset "brandDZ " with Translation 52
3.8 Results of dataset "Djezzy " Without Translation 53
3.9 Results of dataset "brandDZ "Without Translation 53
3.10 Fusion with "bilstm " 55
3.11 Fusion with bilstm 55
3.12 Fusion with bilstm 55
3.13 Fusion with CNN 56
3.14 Fusion with CNN 56

Abbreviations

NLP	Natural Language Processing.
SA	Sentiment analysis .
ML	Machine Learning.
NB	Naive Bayes.
SVM	Support Vector Machine.
TF	Term Frequency.
BoW	Bag of Words.
LSTM	Long Short Term Memory.
CNN	Convolution Neural Network.
BERT	Bidirectional Encoder Representation from Transforms.
RNN	Recurrent Neural Network.
RF	Random Forest.
BiLSTM	Bidirectional Long Short Term Memory.
GRU	Gated Recurrent Unit.

General introduction

Over the past decade, the field of artificial intelligence has witnessed rapid advancements, particularly in the branch of Natural Language Processing (NLP), which has become a cornerstone in the development of intelligent applications capable of understanding and interpreting human language in a way that closely resembles human comprehension. Among the vital applications of this field is sentiment analysis, which aims to identify the emotional attitudes and impressions embedded in written or spoken texts, and classify them as positive, negative, or neutral sentiments.

This application has found wide usage in various domains, including digital marketing, service quality monitoring, opinion analysis on social media, and even in the political arena to understand public orientations. Despite the significant successes achieved in sentiment analysis for formal languages such as Modern Standard Arabic and English, the processing of dialects—especially the Algerian dialect—still poses a major challenge in this field.

This difficulty stems from the unique characteristics of the Algerian dialect, which is distinguished by its multilingual roots, incorporating vocabulary and structures from French, Amazigh, and Modern Standard Arabic. Additionally, the lack of standardized and agreed-upon writing conventions results in significant variability in text representation. The scarcity of structured linguistic resources, such as labeled datasets and dictionaries specific to Algerian dialect, also limits researchers' ability to develop accurate and efficient models.

The importance of this work lies in addressing the challenges of sentiment analysis in texts written in Arabic dialects, particularly Algerian Arabic. To this end, the project proposes an innovative hybrid approach based on late fusion of representations extracted from two advanced language models: DziriBERT, specifically trained on Algerian dialectal texts, and MARBERT, designed to handle the diversity of Arabic dialects. These representations are integrated within a BiLSTM-based classification architecture, which is well-suited for capturing contextual dependencies in text. This setup aims to enhance the model's ability to accurately distinguish between different sentiment categories in a linguistically non-standard and complex environment.
Empirical experiments have demonstrated the superiority of this approach over many traditional and modern models, opening promising prospects for the development of

more specialized and reliable sentiment analysis tools for the Algerian dialect.
This work is structured into three chapters.
The first chapter introduces key concepts in Natural Language Processing (NLP), focusing on the Arabic language and the unique challenges of processing the Algerian dialect.
The second chapter reviews existing research in sentiment analysis, covering both traditional and deep learning methods, with attention to dialectal Arabic advancements.
The third chapter outlines the proposed methodology, detailing data preparation, model design, and experimental results.

.

Chapter 1: Background

1.1 Introduction

Since machines can not handle human natural language, appears the necessity for Natural language processing techniques. In this chapter we present the basic theoretical principles of Natural language processing. We try to clarify and cover the subject from different sides by defining every related aspect to the topic from the overall to the detailed. Beginning with Natural language processing definition, branches and Applications followed by a an overview about Arabic languages and it's categories and the difficulties of applying NLP techniques on the Arabic text.

1.2 Natural Language processing

1.2.1 Definition

Natural Language Processing (NLP) is a subfield of Artificial Intelligence (AI) that draws from computational linguistics and focuses on modeling human language. NLP enables computers to understand, interpret, and generate human language in both spoken and written forms. This involves converting natural language into structured data, allowing machines to perform tasks such as speech recognition, sentiment analysis, and the generation of questions and responses.

NLP systems are often viewed as pipelines, consisting of multiple logical stages of data processing. Natural language—such as English or Arabic—is the way humans naturally communicate, having evolved through use and repetition. Unlike programming languages, natural language can be spoken or written and is often ambiguous and complex. Today, machines are increasingly capable of understanding and interacting using natural language, as seen in technologies like Amazon's Alexa, Apple's Siri, and Google Assistant.

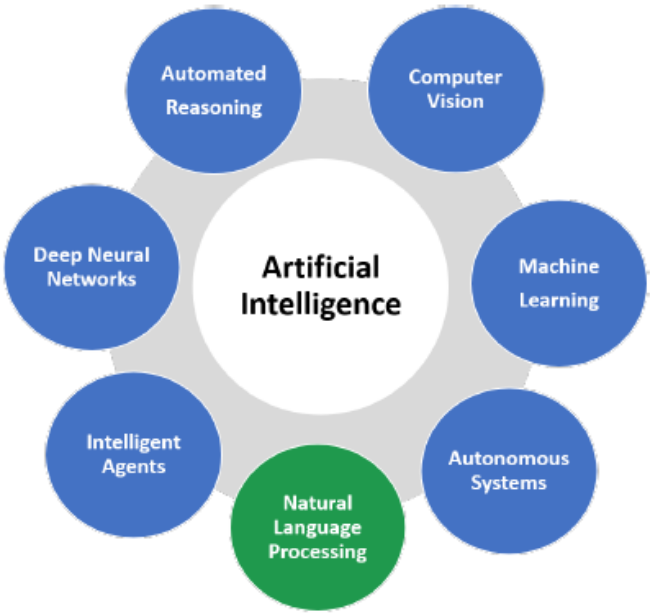


Figure 1.1: Position of NLP in AI [1].

Natural language processing exists since 1940s, at that time NLP systems were implemented manually with code as a set of rules. Since 1990s most of the NLP researches are relied on machine learning. This latter provides automatic learning by using a huge amount of examples to build systems can deal with NLP, yet the biggest advantage of Machine learning for NLP is the accuracy [1].

1.2.2 Historique du NLP

After World War II, early attempts at machine translation—such as the 1954 Georgetown experiment—relied on simple grammatical rules, but proved limited in flexibility. Alan Turing introduced the concept of the "universal machine" and the Turing Test. Meanwhile, Noam Chomsky challenged rule-based approaches, laying the foundation for formal language theory.

In the 1960s and 70s, symbolic NLP emerged, using more advanced rule-based systems. Projects like ELIZA simulated human conversation. During the 1980s and 90s, researchers like Levesque introduced ontologies and began exploring statistical methods.

With the rise of the internet in the 1990s and 2000s, probabilistic models like naïve Bayes and Support Vector Machines (SVMs) were combined with hand-crafted rules, notably by search engines like Google. Tools such as NLTK and spaCy made it easier to develop more advanced NLP systems.

From the 2000s onward, deep learning revolutionized NLP through word embeddings and neural networks. Models like RNNs, CNNs, and LSTMs became common

for processing text sequences. The introduction of attention mechanisms—especially through the landmark paper Attention is All You Need (2017)—led to the development of Transformers, which drastically improved performance.

In 2018, Google introduced BERT, a bidirectional Transformer model, later adapted into French versions like CamemBERT and FlauBERT. Since then, increasingly powerful models have emerged, including those by OpenAI such as Ada, Curie, Davinci, and ChatGPT [2].

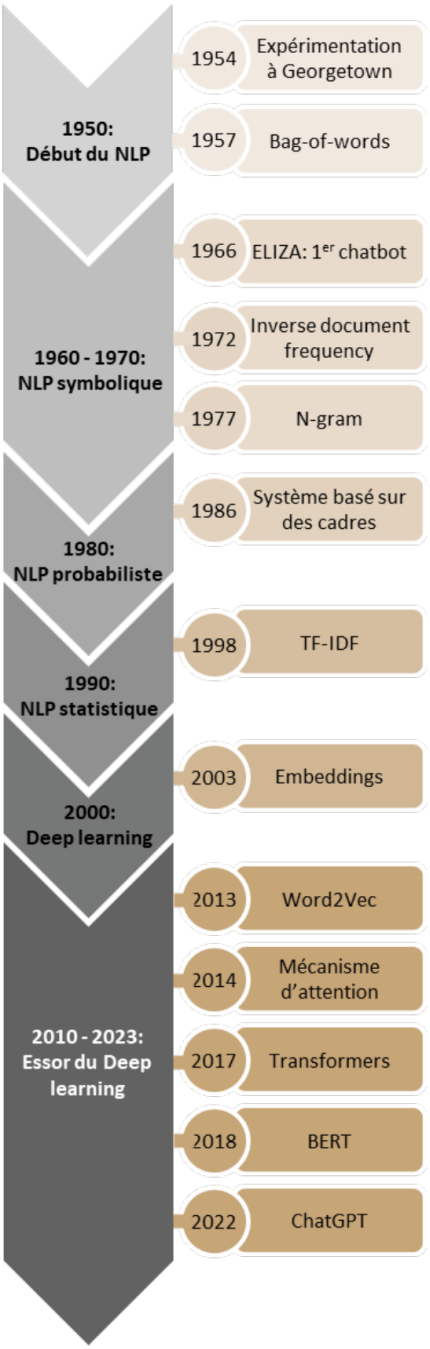


Figure 1.2: Historique du NLP [2].

1.2.3 NLP branches

We can define two aspects for NLP:

Natural Language Understanding (NLU) is a subfield of Natural Language Processing (NLP) that analyzes components of speech to help computers understand and interpret human language. Through machine learning, NLU models continuously improve by learning to identify syntax, context, linguistic patterns, specific meanings, sentiment, and user intent. This capability enables seamless human-computer interaction [3].

Natural Language Generation (NLG) , on the other hand, involves the use of artificial intelligence to automatically produce written or spoken content from structured data. This allows chatbots to access internal or third-party data sources and generate meaningful responses based on the retrieved information [3].

1.2.4 NLP in action



Figure 1.3: NLP in action [4].

Text input and data collection

In Natural Language Processing (NLP), text input and data collection are fundamental steps, as they provide the raw material needed to train language models. This data is gathered from various sources such as social media, forums, emails, and documents. Data collection goes beyond simply inputting text—it involves structured methods like web scraping, database querying, and extracting user-generated content from online platforms. The quality and diversity of the collected data directly impact the effectiveness of NLP models, enabling them to understand dialects, slang, and specialized vocabulary. A thorough and diverse data-gathering process improves the performance of NLP applications such as chatbots, translation systems, and sentiment analysis tools [4].

Text preprocessing

Text preprocessing is a vital step in NLP that prepares raw text for effective analysis and modeling. It begins with tokenization, which breaks the text into smaller units like words or phrases. This helps NLP algorithms better process the text. The next step is removing stopwords—common words like ”the,” ”is,” and ”and”—to focus on more meaningful content. Additional techniques include stemming and lemmatization, which reduce words to their base or dictionary forms, helping standardize and simplify the text.

Other preprocessing tasks include converting text to lowercase, removing punctuation, special characters, and extra spaces, all of which contribute to consistency and cleaner data. These steps improve the quality of the input, allowing NLP models to perform tasks such as classification, sentiment analysis, and translation more accurately and efficiently [4].

Text representation

Text representation in NLP is the process of converting raw text into structured, numerical formats that machine learning models can interpret. One basic method is Bag of Words (BoW), which represents text by word counts or frequencies, ignoring word order. While simple and efficient, BoW often results in high-dimensional and sparse data.

An improvement over BoW is TF-IDF (Term Frequency-Inverse Document Frequency), which assigns weights to words based on their importance within a document and across a corpus, helping highlight meaningful terms while downplaying common ones.

More advanced methods include word embeddings like Word2Vec and GloVe, which map words into continuous vector spaces and capture semantic relationships between

them. These vectors allow models to understand word meanings based on usage.

The most recent advancement is contextual embeddings from models like BERT, which generate dynamic word vectors based on the context in which words appear. This allows for a deeper understanding of language and improves performance in tasks like named entity recognition, question answering, and machine translation [4].

Feature selection

Feature selection in NLP is essential for improving the performance and efficiency of machine learning models by identifying the most relevant features from text data. It begins with filter-based methods like chi-square tests and mutual information, which rank features independently using statistical measures.

Next, wrapper-based methods evaluate different subsets of features during model training to find the best combination based on performance metrics such as accuracy or F1 score. While more accurate, these methods are computationally intensive.

Embedded methods combine the advantages of both filter and wrapper techniques by integrating feature selection directly into the model. Examples include LASSO regression and decision trees, which automatically reduce or eliminate less important features during training.

By focusing on the most meaningful features, these methods reduce noise and computational load, enhancing model accuracy in tasks like sentiment analysis, topic modeling, and machine translation [4].

Model selection training

Model selection and training are critical steps in NLP that greatly influence the success of tasks like classification, regression, or sequence prediction. The process begins with choosing an appropriate model based on the task's complexity and data characteristics. Simpler models like logistic regression may suffice for basic tasks, while advanced models like BERT or GPT are better suited for complex, context-rich applications.

Once a model is selected, training involves feeding it preprocessed text data to learn patterns and relationships. The dataset is typically split into training, validation, and test sets to ensure robust learning and generalization. During training, the model uses a loss function to measure error and applies gradient descent to optimize performance.

An essential part of training is hyperparameter tuning, which adjusts elements like learning rate, batch size, or network depth to enhance performance. Techniques like grid search or random search are used to find the best settings.

Finally, model evaluation uses metrics such as accuracy, precision, recall, and F1 score to assess how well the model performs on unseen data. A careful approach to model selection, training, tuning, and evaluation helps build strong NLP systems that

deliver accurate and meaningful results across various applications [4].

Model deployment inference

Model deployment and inference are the final stages in an NLP project, where the trained model is moved into a production environment to handle real-time data and generate predictions. Deployment involves setting up infrastructure like servers and APIs to ensure the model is accessible, scalable, and can perform efficiently in real-world applications.

Once deployed, the model performs inference, analyzing new data to produce outputs such as classifications, translations, or sentiment scores. This enables practical use in tasks like customer support automation, content recommendation, or real-time language translation.

To maintain performance, continuous monitoring is essential. Key metrics such as response time and accuracy are tracked to detect issues and identify areas for improvement. Over time, model updates and maintenance—including retraining with new data—are necessary to adapt to evolving needs and ensure the model remains accurate and effective.

By carefully managing deployment, inference, and ongoing updates, organizations can fully leverage NLP models to enhance their applications and achieve strategic goals [4].

Evaluation optimization

Assessment and refinement are essential steps in developing effective NLP models, ensuring they meet performance goals and function reliably. Evaluation involves testing the model using metrics such as accuracy, precision, recall, F1 score, and AUC-ROC to measure how well it performs on tasks like classification or sentiment analysis. This is typically done using a test set the model hasn't seen before to fairly assess its generalization ability.

Following evaluation, optimization aims to boost model performance by tuning hyperparameters (e.g., learning rate, batch size, number of layers) using techniques like grid search, random search, or Bayesian optimization. Structural changes or regularization may also be applied to prevent overfitting and improve generalization.

Another key part of refinement is feature engineering, which enhances input quality by selecting or generating more informative features and reducing irrelevant ones, thus improving prediction accuracy.

Finally, continuous monitoring and retraining are crucial once the model is deployed, ensuring it stays accurate and effective as data changes over time. By combining thorough evaluation, ongoing optimization, and active maintenance, organizations

can build robust NLP models that deliver reliable and actionable insights [4].

Iteration improvements

Iteration and improvement are key to building robust NLP models, enabling continuous refinement through repeated cycles of training, evaluation, and adjustment. Each iteration allows for testing changes in architecture, hyperparameters, or feature engineering to identify what enhances performance.

Improvements are based on insights gained from experiments, helping correct weaknesses and meet evolving needs. This may involve adopting advanced techniques like transformer models or contextual embeddings to better handle language complexity and real-world challenges.

The iterative process also includes performance monitoring and user feedback to detect issues such as performance drops or data shifts. Insights from end-users or experts help tailor the model to practical needs, ensuring ongoing relevance.

By continuously learning from results and feedback, NLP models can adapt to new challenges, maintain high accuracy, and stay effective in a variety of applications. This feedback-driven, iterative approach ensures the model remains reliable, impactful, and aligned with user expectations [4].

1.2.5 NLP Applications

Below are some applications of NLP [2]:

- **Machine Translation:** Machine translation systems, such as Google Translate or DeepL, use NLP to translate texts from one language to another. Some translators are also capable of automatically detecting the input language.
- **Automatic Text Generation:** Large Language Models (LLMs) can generate coherent and contextually relevant text based on a given prompt, opening possibilities for creative writing, blog content, etc. Text generation is also used in virtual assistants and chatbots.
- **Sentiment Analysis:** This is the process of classifying the emotional intent of a text. Typically, the input of a sentiment classification model is a piece of text, and the output is the probability that the expressed sentiment is positive, negative, or neutral. Sentiment analysis is used, for example, to classify customer reviews on various online platforms.
- **Topic Modeling:** This is a text analysis technique used to discover patterns or latent themes within a corpus of documents. The goal of topic modeling is to automatically extract topics or themes from large sets of unlabeled texts, without a human having to predefine these themes. This is especially useful for recommending documents based on an initial document and for identifying trends.

- **Text Classification:** The most common use case is spam detection. This is a widespread binary classification problem in the NLP field, aimed at classifying emails as spam or not. Spam detectors take as input the text of an email as well as various other subtexts such as the title and the sender's name. They aim to determine the probability that the email is spam. Email service providers like Gmail use such models to improve user experience by detecting unsolicited and unwanted emails and moving them to a designated spam folder.
- **Grammatical Error Correction:** Grammar correction models encode grammatical rules to correct the grammar of a text. This is mainly a sequence-to-sequence task, in which a model is trained with an ungrammatical sentence as input and a correct sentence as output. Online grammar checkers like Grammarly and word processing systems like Microsoft Word use these systems to offer a better writing experience to their users.
- **Text Summarization:** This consists of shortening a text to highlight the most relevant information.
- **Speech Recognition:** These are algorithms that can identify a speaker's voice, convert spoken words into text, and interpret the underlying meaning. Apple's Siri and Amazon's Alexa are tools that use NLP to listen to user requests and find responses. Many companies also integrate these models into their Customer Relationship Management (CRM) systems to improve the customer experience [2].

1.2.6 NLP advancements

Natural Language Processing (NLP) has a wide range of applications across numerous fields. Search engines like Google, Bing, and Yahoo use NLP to interpret user queries, index content, and rank results to deliver the most relevant information. Sentiment analysis is employed to detect emotions or opinions in text—such as product reviews or social media posts—helping businesses better understand customer feedback.

Chatbots and virtual assistants like Siri, Alexa, Google Assistant, and Cortana rely on NLP to understand and respond to user requests in natural language. Machine translation services such as Google Translate, DeepL, and Microsoft Translator utilize NLP to translate text between languages accurately.

Text classification, another important application, involves categorizing text into predefined categories and is widely used in spam filtering, document classification, and automated tagging. Information extraction helps identify and extract key data from text, such as through Named Entity Recognition (NER). Text summarization tools create concise versions of longer texts, which is particularly useful for news articles and reports.

Plagiarism detection tools like Turnitin and Grammarly use NLP to identify copied

or paraphrased content in academic and professional documents. In healthcare, NLP is used to analyze medical records and scientific literature for tasks like disease classification and clinical data extraction, supporting automated medical assistants and enhanced patient care.

In the financial and business sectors, NLP assists in analyzing news articles, market reports, and social media trends to forecast market movements. Recommendation systems on platforms such as Netflix, Amazon, and YouTube leverage NLP to understand user behavior and deliver personalized content suggestions. NLP also plays a key role in fake news detection by identifying and classifying misleading or false content.

Language learning applications like Duolingo use NLP to create customized exercises and provide personalized feedback to learners. Additionally, in cybersecurity, NLP is used to analyze logs and communications to identify threats and malicious activities. These examples highlight the powerful and versatile role of NLP in transforming and enhancing various industries [4].

1.3 Arabic language

Arabic language [5] is a Semitic language spoken by approximately 420 million people. It is the official language in approximately 22 countries. Arabic is a broad term that encompasses three distinct groups: classic Arabic, modern Standard Arabic and dialectal Arabic.

1.3.1 Classic Arabic (CA)

is primarily defined as the Arabic used in the Qur'an and the earliest literature from the Arabian peninsula, but it also serves as the foundation for much literature up to the present day.

1.3.2 Modern Standard Arabic (MSA)

(Alfus'ha in Arabic) is the variety of Arabic that has been retained as the official language in all Arab countries and as a common language. It is basically a modernized version of classical Arabic. Standard Arabic is not the first language most Algerians pick up. MSA is learned at school and through exposure to formal broadcast programs (such as the daily news), religious practice, and newspapers.

1.3.3 Dialectal Arabic

Also known as colloquial Arabic or vernaculars are spoken dialects of Arabic. They are not written, unlike classical Arabic and MSA. These dialects have a hybrid form with

numerous variations. They are influenced by both ancient indigenous languages and European languages such as French, Spanish, English, and Italian. The differences in these dialects of spoken Arabic throughout the Arab world can be significant enough to render them incomprehensible to one another. As a result of the significant differences between dialects, we can regard them as distinct languages based on the geographical location in which they are practiced. As a result, the majority of the literature describes Arabic dialects from an east-west perspective [6].

1.4 Algerian dialect (AD)

In Algeria, as in other places, spoken Arabic differs from written Arabic. Algerian dialect has a vocabulary inspired by Arabic, but the original words have been phonologically altered, with many new words and phrases. loanwords from French, Turkish, and Spanish. Like all Arabic dialects, Algerian Arabic has dropped the case endings of the written language.

Algerian dialect is not used in schools, on television, or in newspapers, which typically use standard Arabic, English, or French, but is more likely to be heard in songs, if not only in songs. Algerian houses and streets Algerian Arabic is widely spoken. The vast majority of Algerians use it every day [6].

1.4.1 Characteristics of Algerian Dialect

Algerian Arabic is the language spoken in Algeria that is distinguished by a lack of standardization and writing resources. It is regarded as a language with insufficient resources. It is not the same as Modern Standard Arabic in any aspect of linguistic representation, including morphology, syntax, lexicon, and phonology.
In this section, we present their most important characteristics [7]:

Lexical characteristics : Like other Arabic dialects, the Algerian dialect has been impacted over centuries by languages such as Berber, Spanish, French, Italian, and Turkish.

Morphological characteristics : Algerian dialect also has different morphological aspects, consisting essentially of a simplification of some inflections and the inclusion of new clitics as follows:
- The casual endings in nouns and verbs mood are lost;
- The indicative mood is the one that is used as default unlike the other moods that are not used;

- The dual and the feminine plural vanished. And they have been absorbed into the masculine plural;

- The first and second person of the singular form are conjugated in the same way, for example, ;‘thank you/I ‘škart

Orthographical characteristics : There are two causes for the orthographic variance in Arabic dialect word writing [8]:
-The absence of an orthographic standard due to the lack of codification and standardization of Arabic dialects.
-The phonetic distinctions between the Algerian dialect (ALG) and MSA.

1.4.2 Difficulties of Arabic language processing

NLP plays a key role in the preparation stage for every Arabic language matter. When applying NLP techniques to the Arabic language, several challenges appear beyond what is expected for English. The Arabic language consists of 28 distinct characters that are written from right to left. There are Arabic language features that are inherently challenging for Arabic NLP researchers and developers.These features include the nonconcatenative nature of Arabic morphology, the absence of the orthographic representation of Arabic short vowels from contemporary Arabic texts, and the need for an explicit grammar of MSA that defines linguistic constituency in the absence of case marking.

Another big issue with the Arabic script is that the letters change shape based on their position in the text. The originality of the words is a final challenge; 85 percent of Arabic words were formed from their roots [9]. Important aspects such as anaphoric relations, subjectless sentences, and discourse analysis must be described.

1.5 Conclusion

Humans use natural language to communicate with each other. To achieve the goal of understanding natural language by machines, natural language processing (NLP) techniques were invented and developed.

Since NLP is the pillar domain to any text handling technique, in this chapter, we focused on natural language processing by presenting a brief NLP definition and its branches. In addition, we displayed a set of NLP applications. Then we defined the Arabic language and its groups and also spotted the light on the Algerian dialect. Finally, we mentioned some difficulties withArabic language processing.

Chapter 2: Related work

2.1 Introduction

As communication on the internet keeps growing, social media platforms have also become powerful spaces where people can express themselves and influence millions of other persons. This constant interaction has changed almost everything about life such as social, political, commercial or cultural lives thus requiring sentiment analysis to help understand dynamics in public opinions. In a famous quote from Edward Bulwer-Lytton "The pen is mightier than the sword" he highlights the role played by free speech. Even though this freedom encourages self-expression, there are fears that if not properly used, these channels of communication may affect individual rights and even societal norms.

Sentiment analysis has a special importance in multilingual settings and e-commerce since customer sentiment understanding is vital for informed decisions and effective strategies development.

2.2 Sentiment Analysis

2.2.1 Definition of Sentiment Analysis

Sentiment analysis (SA) , also known as Opinion mining (OM), is the computational study of people's beliefs, attitudes, and feelings concerning a subject. The object may stand in for people, things, or subjects. Reviews are more likely to cover these subjects. You can use either SA or OM interchangeably. They both convey the same notion. On the other hand, several researchers claimed that SA and OM had rather different ideas [10].

Sentiment analysis finds and evaluates the sentiment expressed in a text, whereas opinion mining gathers and examines people's opinions about an object. As a result, the goal of SA is to gather viewpoints, determine the emotions they convey, and categorize their polarity .

2.2.2 Types of Sentiment Analysis

There Are so many types on Sentiment Analysis :

Graded Sentiment Analysis:

There are three main categories in sentiment analysis: positive, negative, and neutral. However, sentiment analysis can also capture a broader range of emotions, such as happiness, sadness, and anger, depending on the context and application needs. For

instance, in e-commerce platforms, customer product reviews reflect sentiment. Consider the review: "This product is great, delivered on time, and I'm very satisfied." The true sentiment or emotion expressed by the customer here is positive or happy [11].

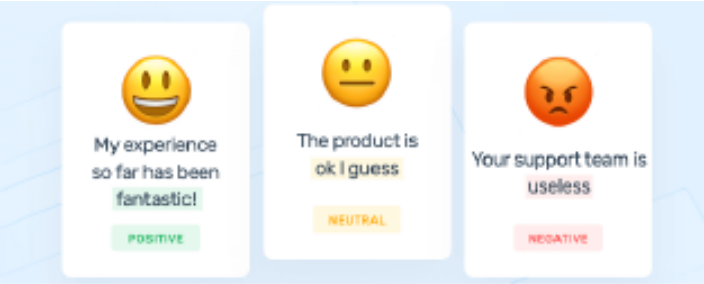


Figure 2.1: Sentiment analysis grades [11].

Emotion Detection :

On the other hand, emotion detection is the process of recognizing different kinds of human emotions in text , images or audios, like anger, happiness, or depression. There are instances when terms like "emotion detection," "affective computing," "emotion analysis," and "emotion identification" are used interchangeably [11].



Figure 2.2: Emotion classes in sentiment analysis[11].

Multilingual sentiment analysis :

The birth of the Internet meant people started voicing their views on all sorts of things like products, services, happenings, and politics through sites, blogs, and social media in myriad languages , So they added another type of sentiment analysis called " Multilingual sentiment analysis". Analyzing opinions in multilingual texts is multilingual sentiment analysis (MSA). This creates hurdles like language differences, cultural nuances, linguistic subtleties, plus resource-constrained tongues lacking quality data. Resource constrained languages suffer from a dearth of wellcurated or suitably tagged data for effective use [12] .

2.2.3 Sentiment Analysis Levels

To apply sentiment analysis we should define the text that will be analyzed in the case of a study considered. Sentiment analysis can occur at generally three levels:

Document level: It determines the overall sentiment of a text (classifying a review as positive, negative, or neutral). It works best when a single person writes the document and expresses only one perspective on a single entity [13].

Sentence level: it determines the polarity of each sentence contained in a text; it usually involves subjectivity classification of the sentence (whether the sentence is objective or subjective) and sentiment classification of subjective sentences (positive or negative) [**?**].

Aspects/ Features level: It performs a more detailed analysis compared to other levels. Based on the idea that an opinion consists of a sentiment and a target, it identifies and extracts the features of an object that have been commented on by the opinion holder and determines whether the opinion is positive, negative, or neutral. This level of analysis enables the differentiation of aspects that are liked or disliked by the authors of the texts, making it easier to identify potential areas for improvement or treatment [13].

Sentiment analysis is extremely useful in social media monitoring because it provides an overview of the public's opinion on certain topics. The ability to extract insights from social web data is a practice that is widely adopted by companies around the world; therefore, the use of sentiment analysis is both broad and powerful

2.3 Sentiment analysis methods

Generally, SA approaches can be divided into four categories: lexicon-based, traditional machine-learning, deep-learning and hybridapproaches . Sentiment analysis approaches are depicted in Fig 2.3 [14] .

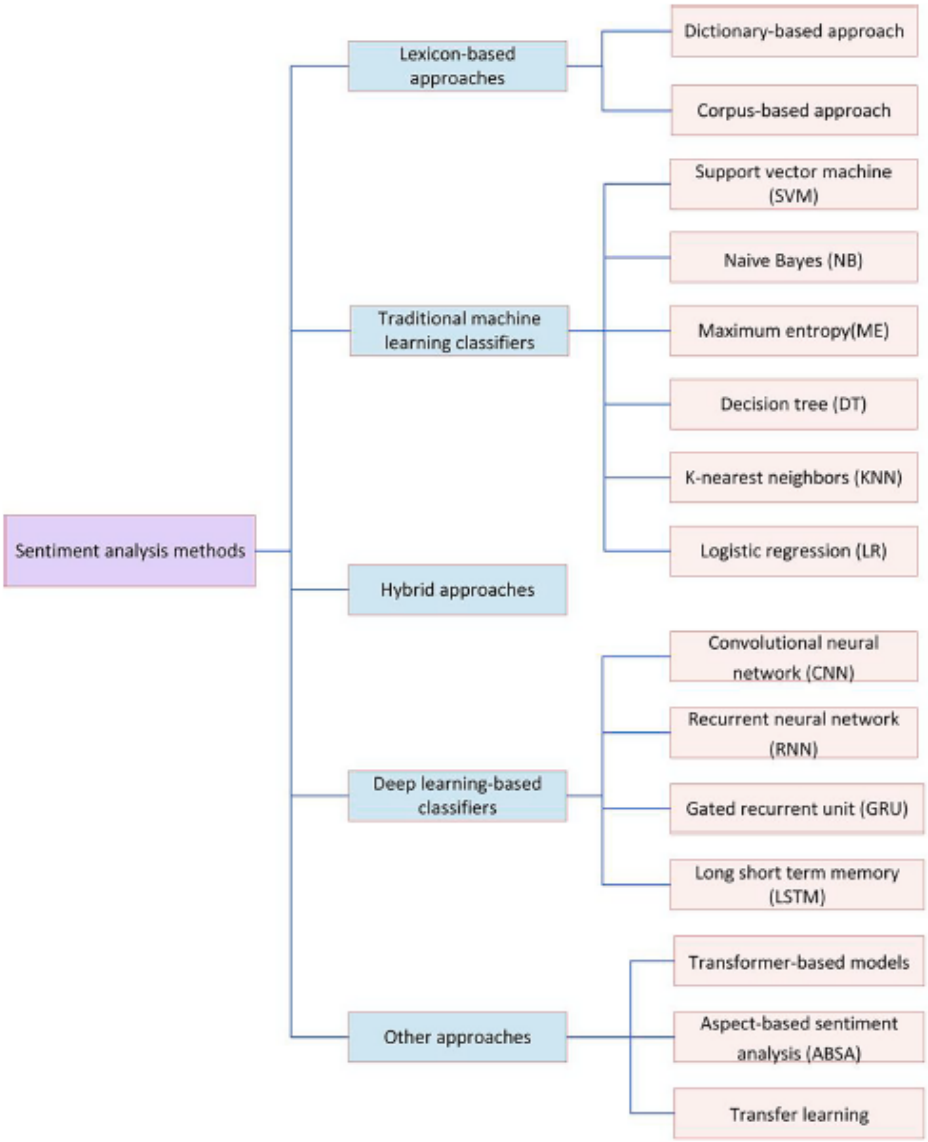


Figure 2.3: Sentiment analysis methods[14]

2.3.1 Lexicon-based approach

Lexicon-Based or Subjective lexicons approach are collection of words where each word has a score indicating the positive, negative, neutral and objective nature of text. In this approach, for a given piece of text, aggregation of scores of subjective words is performed i.e. positive, negative, neutral and objective word scores are summed up separately. In the end there are four scores. Highest score gives the overall polarity of the text [15].

- **Dictionary based approach**

In this approach a set of opinion words are manually collected and a seed list is prepared. Then we search for dictionaries and thesaurus to find synonyms and antonyms of text. The newly found synonyms are added to the seed list. This process continues until no new words are found. Disadvantage: difficulty in finding context or domain oriented opinion words [15].

- **Corpus based approach**

Corpus is collection of writings, often on a specific topic. In this approach, seed list is prepared and is expanded with the help of corpus text.Thus it solves the problem of limited domain oriented text. It can be done in two ways [15]:

Statistical approach: This approach is used to find co-occurrence words in the corpus. Idea is that if the word appears mostly in positive text, then its polarity is positive. If it mostly occur in negative text, then its polarity is negative .

Semantic approach: This approach calculates sentiment values by using the principal of similarity between words. Wordnet can be used for this purpose. Synonyms and antonyms of given word can be found using this and sentiment value can be calculated.

2.3.2 Traditional machin learning classifiers

"Machine Learning is a field of study that gives computers the ability to learn without explicitly being programmed.," according to Arthur Samuel's 1959 definition [16].

In general, machine learning is a relatively new use of artificial intelligence (AI) that gives computers the capacity to autonomously learn from experience and get better at it without explicit programming. The concept is to provide machines with access to data so they can learn on their own. This is predicated on the machines' ability to comprehend the structure of data and transform it into models that humans can comprehend and use [17] .

Methods

Machine learning techniques are often categorized into two groups: Supervised Learning for classification and regression, Unsupervised Learning for clustering . Additionally machine learning methods include reinforcement learning and deep learning are two more machine-learning techniques .

Supervised Learning: The machine learning task of supervised learning involves using sample input-output pairs to train a function that maps an input to an output. From labelled training data, which consists of a collection of training instances, it infers a function. The machine learning algorithms that require external aid are known as supervised algorithms. The train and test datasets are separated from the input dataset. An output variable from the train dataset needs to be categorized or forecasted. Every algorithm picks up some sort of pattern from the training dataset and uses it to predict or classify data from the test dataset [18].

- **Support vector machine (SVM)** Support Vector Machines (SVM) are widely used in sentiment analysis to classify text data based on sentiment, such as reviews, social media posts, or comments. SVM works by finding the hyperplane that best separates different classes in a high-dimensional feature space. In the context of sentiment analysis, these classes typically represent different sentiments such as positive, negative, or neutral [19].

In the study "Sentiment Analysis for Arabizi: Application to Algerian Dialect," the Support Vector Machine (SVM) algorithm played a key role in improving the accuracy of sentiment analysis for texts written in the Algerian dialect using the Latin alphabet (Arabizi). Thanks to its ability to handle high-dimensional data and nonlinear classifications, SVM outperformed other algorithms such as Naive Bayes and Decision Trees in classifying sentiments as positive, negative, or neutral. The results showed that SVM achieved the best performance, with classification accuracy increasing from 76% to 87% after applying preprocessing techniques, confirming its effectiveness in analyzing non-standard texts and achieving more accurate and reliable classifications [20].

- **Naive bayes (NB)** Naïve Bayes is a popular method for sentiment analysis, used to automatically classify texts into positive, negative, or neutral categories. Based on Bayes' theorem, this algorithm assumes that the words in a text are independent of each other, which simplifies the calculation of probabilities. It is often applied to customer reviews, social media comments, or movie reviews. Its main advantage is its speed and efficiency, even with large datasets. However, it can be improved by integrating advanced natural language processing techniques to better capture the context of words [21].

"Sentiment Analysis using Naïve Bayes Algorithm" is an article that discusses the role of the Naïve Bayes algorithm in sentiment analysis, where its success rate ranges between 75% and 90%, depending on data quality and sample size. The algorithm performs well with short and simple texts but may struggle with understanding complex expressions such as sarcasm. Its performance can be improved using techniques like TF-IDF or word embeddings, yet it remains less accurate compared to

more advanced algorithms like SVM or deep neural networks [21].

• **Decision tree (DT)** Decision Trees (DT) are widely used in sentiment analysis to classify texts into categories such as positive, negative, or neutral. For instance, a study applied DT algorithms to analyze movie reviews on IMDB and Twitter data, achieving classification rates of 94.00 % and 82.76 %, respectively [22].

Another research compared several supervised learning algorithms, including Decision Trees, for sentiment classification in product reviews, demonstrating the effectiveness of these models in detecting expressed sentiments [23].

These studies highlight the utility of Decision Trees in sentiment analysis, offering clear interpretation and the ability to handle non-linear relationships between textual data features.

• **K- nearest neighbors (KNN)** The k-Nearest Neighbors (KNN) algorithm is a widely used method in sentiment analysis for classifying textual data, such as product reviews, social media posts, and user comments. KNN operates on the principle that similar data points are close to each other in the feature space.

In sentiment analysis, the process begins with feature extraction from the text, often utilizing techniques like n-grams. These features are then used to train the KNN model, which learns to associate specific feature sets with particular sentiments—positive, negative, or neutral.

A study titled "Sentiment Analysis Approach Based on N-gram and KNN Classifier" explores this methodology by combining n-gram feature extraction with KNN classification to analyze sentiments in Twitter data. The experimental results demonstrate that this approach achieves good performance in terms of precision, recall, and accuracy, outperforming other methods like the Support Vector Machine (SVM) classifier in certain cases [24] .

• **Logistic regression (LR)** Logistic regression is commonly used in sentiment analysis to classify texts into positive or negative categories. For example, one study compared several machine learning algorithms, including logistic regression, to analyze tweets in English, Filipino, and Taglish, and found that logistic regression had the best accuracy with a rate of 81% [25].

Additionally, another research applied regularized logistic regression to online reviews, achieving a true positive rate of 94.9 in sentiment classification [26] .

These studies demonstrate the effectiveness of logistic regression in sentiment analysis, especially when combined with appropriate text data preprocessing techniques.

Unsupervised Learning : In unsupervised learning, the training data is unknown and unlabeled, so the learning algorithm is left to find commonalities between its inputs data. Without the aspect of known data, the input cannot be guided to the algorithm, which is where the unsupervised term originates. Unsupervised learning goal is discovering and modelling the hidden patterns, or the underlying structure in the given input data, so as to learn about the data, which is fed to the learning algorithm and used to train the model, that help to find features which can be useful for categorization. The trained model tries to search for a pattern and give the desired response. As unlabeled data are more numerous and obtainable than labeled data, machine learning methods that facilitate unsupervised learning are particularly valuable. Otherwise, we cannot get precise information, and the results are less accurate, because the input data is not known and unlabeled by humans, the machine requires doing this itself. Unsupervised learning is commonly used for transactional and preprocessing data, during exploratory analysis, or to pre-train supervised learning algorithms [27].

Reinforcement Learning : Richard in [28], defined reinforcement learning as the learning of a mapping from situations to actions, to maximize a scalar reward or reinforcement signal. The learner is not told which action to take, as in the most forms of machine learning, but instead must discover which action produces the highest reward by trying them. Reinforcement Learning algorithms is a zone of Machine Learning inspired by behaviorist psychology and biological learning systems . It deals with learning via interaction and feedback; it solves tasks with trial and error by acting in an environment and receiving rewards for it. Reinforcement learning contrasts with other machine learning approaches in that the algorithm is not explicitly told how to perform a task, but works through the problem on its own. It is about making decisions sequentially, where the output depends on the state of the current input, and the next input depends on the output of the previous input. Therefore, we do not give labels to each decision but we give labels to sequence of dependent decisions .

2.3.3 Deep learning-based classifiers

By using a hierarchy of concepts, deep learning is a type of machine learning that lets computers learn from experience and comprehend the world. The computer taught itself how to handle and absorb data [29] .

It operates on a network of synthetic neurons that are modeled after the human brain. The neurons in this network are arranged in tens or even hundreds of layers, each of which receives and processes data from the layer above it [29].

Neural networks and deep learning differ in the depth of the model; deep learning is a term for intricate neural networks. It has to do with feature extraction and trans-

formation, which try to establish a connection between inputs and the corresponding neural responses that the brain produces. Elaborate patterns of how information can flow across the model confer complexity. The most popular deep learning algorithms are : Convolutional Neural Network (CNN) ,Recurrent Neural Networks (RNN) , Long short term memory Networks (LSTM)s [29].

- **Concolutional neural network (CNN)** Convolutional Neural Networks (CNN) are widely used in sentiment analysis to extract local features from texts and improve the accuracy of emotion classification. For example, one study demonstrated the effectiveness of CNNs by combining them with Bi-LSTM models to analyze customer reviews, achieving significant accuracy in sentiment detection [30].

  Another research applied CNNs to tweet data from the Weibo platform, achieving an F1 macro-average score of about 0.73, highlighting the robustness of this approach in sentiment classification [31].

  These studies show that CNNs, by capturing local text structures effectively, play a key role in enhancing the performance of sentiment analysis systems.

- **Recurrent neural network (RNN)** Recurrent Neural Networks (RNNs) are highly effective tools for sentiment analysis due to their ability to process sequential data and understand context in texts. Unlike traditional models, RNNs have recurrent connections that allow past information to influence current outputs, making them well-suited for handling text, where the meaning of a word often depends on previous words in a sentence.

  For example, in the study **"Character Gated Recurrent Neural Networks for Arabic Sentiment Analysis"**, an improved RNN model based on character-level representations was used for sentiment analysis in Arabic texts. This model demonstrated high efficiency in understanding the complexities of the Arabic language and handling challenges such as variations in word forms and dialect diversity. It was trained on a large dataset of labeled Arabic texts, achieving high accuracy in sentiment classification [32].

  This example highlights the importance of using RNNs for sentiment analysis, especially in languages with complex structures. By leveraging RNNs, businesses and organizations can gain precise insights into customer interactions and opinions, leading to better decision-making.

- **Long short term memory (LSTM)** Long Short-Term Memory (LSTM) networks are widely used in sentiment analysis to process text sequences and identify expressed emotions. For instance, one study applied LSTM to analyze movie reviews on IMDB, achieving an accuracy of 86.1 % in sentiment classification [33].

Additionally, another research used LSTM to analyze tweets, achieving 90 % accuracy in detecting expressed sentiments [34] .

These studies demonstrate the effectiveness of LSTM in sentiment analysis, particularly when combined with appropriate text data preprocessing techniques.

2.3.4 Tansformer-based models

BERT or Bi-directional Encoder Representations from Transformers was proposed by Jacob Devlin in 2018.It is the first deeply bidirectional, unsupervised NLP model that makes use of transfer learning method.

BERT is a pre-trained, bi-directional language model, capable of giving consideration to the context of tokens that appear before and after a selected token. In simple words we can say, BERT can read a text in left to right and right to left direction, giving each word a proper meaning by keeping in mind both contexts, in one single iteration. In addition to being deeply bidirectional, it also has a unique feature that it can be easily applied to perform any new NLP task with slightest changes to the pre-training model. These slightest changes often consist of applying an additional output layer to the pre-trained model [35].

Several variants of BERT have been developed to optimize its performance or adapt the model to specific needs:

RoBERTa (Robustly Optimized BERT Pretraining Approach): This version optimizes BERT training by removing the task of predicting the next sentence and using dynamic masking. RoBERTa is trained on larger data sets and with larger batches, leading to better performance on various natural language processing tasks.

ALBERT (A Lite BERT): ALBERT reduces the size of the BERT model by sharing parameters across layers and factoring embeddings. This approach reduces the number of parameters while maintaining comparable performance, making the model lighter and faster to train.

DistilBERT: This version is a distillation of BERT that reduces the model to 60% of its original size while retaining 95% of its performance. DistilBERT is faster and less resource intensive, making it suitable for applications requiring increased efficiency.

ELECTRA: ELECTRA introduces a new training method where a small model generates plausible substitutions for masked tokens, and a larger model must identify these replaced tokens. This approach improves training efficiency compared to traditional masking used in BERT.

DeBERTa (Decoding-enhanced BERT with Disentangled Attention): DeBERTa modifies the architecture of BERT by separating positional and token encodings throughout the attention mechanism. This approach allows for better capture of the relationships between words and their positions in a sentence, thereby improving contextual understanding.

These variants demonstrate the continued evolution of natural language processing models, each providing specific improvements to address various performance and efficiency needs.

2.4 Sentiment Analysis Domain of Application

One of the most important needs of businesses and organizations in the real world is to find and analyze consumer or public opinions about their products and services . Knowing the opinions of existing users regarding a specific product is also interesting for individual consumers. Sentiment analysis paves the way to several and interesting applications, in almost every possible domain. Other projects have as a long-term goal the clarification of politicians' positions, such as what public figures support or oppose, to enhance the quality of information that voters have access to. figure [2.4] shows some of real world applications of sentiment analysis [36].



Figure 2.4: Some Applications of Sentiment Analysis [36].

2.5 The Challenges of Sentiment Analysis :

Sentiment analysis faces several challenges, such as high computational costs, informal writing styles, and linguistic variability. In this study, we explore the common issues that emerge when analyzing particular forms of sentiment structures. Key obstacles in sentiment analysis include:

- **Language Issue :** In Sentiment Analysis, English is widely used because of the abundance of resources like lexicons, dictionaries, and corpora. However, there is growing interest among researchers in applying Sentiment Analysis to other languages, such as Arabic, Chinese, and German. As a result, they face challenges in developing essential resources, including lexicons, dictionaries, and corpora, for these languages [37].

- **Fake Opinion :** Often referred to as a fake review, it describes fraudulent or deceptive evaluations intended to mislead readers or customers by providing dishonest negative or positive opinions about a product or service, often aiming to harm its reputation. Such spam messages undermine the effectiveness of sentiment analysis across various domains. Despite this societal challenge, the field of opinion mining has witnessed significant progress in sentiment analysis (SA) [37].

- **Neutral reports of events:** Before describing events or situations, the speaker must express their emotional state, as it is unclear whether the account is meant to reflect an emotional bias or to convey a specific emotional condition (such as happiness, anger, cheerfulness, or sadness) [38].

- **Sarcasm :** Sarcasm is a common phenomenon on social media and presents challenges for analysis, not only for automated systems but also for human interpretation. Its influence on sentiment is considerable, yet it is often overlooked in social media studies due to its inherent complexity. Although some systems have been developed to detect sarcasm, limited research has focused on evaluating its impact on sentiment in tweets and incorporating it into automated sentiment analysis tools [39].

2.6 E-Commerce and Recommendation Systems

2.6.1 Overview and Definition recommendation Systems

A recommender system is largely used to evaluate consumers' interests, habits, and preferences while selecting a certain product or service. This information comes from a variety of sources, including the user's buying history . These data are then used to estimate if the user would purchase the same product or a comparable product variety in the future. Filtering technology is used to improve the algorithm in this regard . When a product is purchased, it is given a ranking, and data are processed by analyzing the ranked products and the user list [9]. Some algorithms assign weights to products, and then select the highest-ranked or most-weighted products from the list to recommend to the consumer [40].

Recommendation systems were developed to address the issue of information overload in a variety of fields, including e-commerce, entertainment, and content platforms. Users frequently demand assistance accessing large catalogs of products, movies, music, articles, and other content. The goal is to help people discover relevant and tailored items or material based on their choices, interests, and previous behavior. Recommendation systems have the capability to generate personalized suggestions by analyzing individual user preferences through data such as browsing history, purchase history, reviews, and interactions [40].

Recommender systems strive to provide tailored recommendations for products or services, enriching decision-making in the expansive landscape of internet information. Across three fundamental domains—business, government, and education—a plethora of systems has emerged, encompassing eight categories: e-government, e-business, ecommerce/ e-shopping, e- library, e-learning, e-tourism, e resource services, and e-group activities [10]. Particularly in e-commerce, these systems often propose additional products for customers to explore from a vast array, with filtering methods enhancing the presentation of personalized options [40].

2.6.2 Role of Sentiment Analysis in Enhancing Recommendation Systems

Sentiment analysis plays a crucial role in enhancing recommendation systems by leveraging users' opinions and emotions expressed in reviews and comments. Traditionally, recommendation systems rely on explicit data such as product ratings. However, these ratings are often insufficient or absent. Integrating sentiment analysis helps overcome these limitations by providing implicit information about users' preferences [41].

By analyzing user-generated texts, such as reviews and comments, systems can

extract positive, negative, or neutral sentiments associated with specific product or service features. This approach provides a more detailed understanding of users' preferences, allowing for more precise and personalized recommendations. For example, a user may give a high rating to a restaurant while expressing concerns about the service in their review. Sentiment analysis detects these nuances and adjusts recommendations accordingly.

Studies have shown that integrating sentiment analysis improves the performance of recommendation systems. For instance, research published in Sensors proposed a combined approach using sentiment analysis and collaborative filtering, demonstrating a significant improvement in recommendations through deep learning models analyzing user sentiments [42].

Additionally, another study published in Future Internet explored the use of Bi-LSTM models for sentiment analysis integrated into collaborative filtering techniques, showing a notable improvement in recommendation accuracy in the e-commerce sector [43] .

2.7 Comparison study.

2.7.1 Case of Arabic Algerian dialect (Arabic Script)

In 2024, researcher Amin Abdedaiem and his team introduced FASSILA, the first publicly available corpus designed specifically for fake news detection and sentiment analysis in the Algerian Arabic dialect (AD) written in Arabic script. The dataset contains 10,087 sentences and over 19,497 unique words, sourced from social media platforms such as Facebook and YouTube, as well as manually created, translated, and GPT-4–paraphrased sentences. Preprocessing steps included removing special characters, Latin-script words, and normalizing all text into Arabic script. The sentences were categorized by topic (e.g., politics, health, tourism) and annotated by native speakers for truthfulness (real/fake) and sentiment (positive, negative, neutral), achieving a Cohen's Kappa score of 0.6488 for inter-annotator agreement on sentiment labels.

The researchers evaluated both transformer-based models (AraBERTv02, MARBERTv2, DziriBERT) and traditional machine learning classifiers (SVM, Logistic Regression, KNN). MARBERTv2 achieved the best accuracy for fake news detection (accuracy: 0.7898), while AraBERTv02 led in sentiment analysis (accuracy: 0.8392, F1-score: 0.8035). This work highlights the critical role of dialect-specific resources in improving NLP performance in low-resource language settings [44].

In 2021, Imane Guellil and other researchers introduced an innovative approach to sentiment analysis for Algerian dialect comments, combining both automatically generated and manually reviewed corpora. The study involved extracting comments from 226 popular Algerian Facebook pages using the Facebook Rest API with MSA/DALG keywords, resulting in 7,926,504 comments after preprocessing steps such as removing duplicates, special characters, and Tatweel characters. Arabizi messages were transliterated into Arabic script using specific rules, and a machine translation system was developed to translate between Arabizi and MSA. The researchers used an automated method to construct a corpus lexicon by leveraging the Glosbe API, translating an English lexicon into DALG and MSA, which resulted in 1,745 curated sentiment terms. For classification, a comparison was made between shallow algorithms (e.g., GaussianNB, LogisticRegression, RandomForest) and deep learning models (e.g., CNN, LSTM, Bi-LSTM), with Word2Vec and fastText used for feature extraction. The best-performing machine learning model achieved an F1 score of 0.89 on the Test-sentialg corpus, while CNN with fastText also reached an F1 score of 0.89, with better results observed on the manually reviewed corpora [45] .

Table 2.1: Case of Arabic Algerian dialect (Arabic Script)

WORK	**Preprocessing**	**feature extraction**	**classification**	**Dataset statistics**	**Evaluation Protocol**	**Evaluation Results**
[44]	- Text cleaning - Translation (Arabizi → arabe) - Normalisation of Arabic text	Fine-tuned embeddings from pre-trained transformer models: • AraBERTv02 • MARBERTv2 • DziriBERT	**- Transformer-based models:** • AraBERTv02 • MARBERTv2 • DziriBERT **- ML classifiers using embeddings:** • SVM • KNN • Logistic Regression (LR) • Decision Tree (DT) • Naïve Bayes (NB)	**FASSILA** (A corpus of 10,087 Algerian Dialect sentences annotated)	80% for training 10% for testing 10% for validation	**AraBERTv02** was the best performer: (83.92% Accuracy) .
[45]	- Text cleaning - Transliteration - Translation (Arabizi → arabe) - Normalisation of Arabic text	- Word2Vec - fastText	**- Classifieurs classiques :** - GaussianNB (GNB) - Logistic Regression (LR) - Random Forest (RF) - SGDClassifier - LinearSVC **- Classifieurs profonds :** - CNN - LSTM - Bi-LSTM	Facebook comments(Corpus constructed automatically +corpus reviewed manually)	80% for training 20% for testing	- **fastText + CNN:** best result (89% accuracy)
[20]	- Text cleaning - Transliteration - Transliteration from Arabizi to Standard Arabic - Normalisation of Arabic text	- TF-IDF - Bag of Words	- Naive Bayes - SVM (Support Vector Machine) - Random Forest	- Comments from Algerian Facebook pages ("Wacht7ass" and "BrandtDZ")	- Cross-validation	- **SVM** achieved the best performance (83% accuracy)

2.7.2 Arabic Algerian Dialect (Latin or Arabic script)

The study by Mazari and Djeffal (2022) presents an in-depth machine and deep learning approach for sentiment analysis of social media comments in the Algerian dialect. The researchers constructed a dataset of 11,760 manually annotated comments and trained Word2Vec models (CBOW and Skip-Gram) using over 466,000 texts to semantically enhance the data. They applied several preprocessing techniques, including transcription to Latin script and phonetic grouping, to address dialectal variability. Six classifiers were tested: SVM, three Naive Bayes variants, CNN, and RNN. The best accuracy (84.21%) was achieved using Multinomial Naive Bayes on transcribed data, while the SVM reached 83.70% when combined with Word2Vec enhancement. Results confirm the efficiency of text transcription and traditional ML models in dialectal sentiment classification [46] .

The study by Abdelli et al. (2019) proposes a supervised approach for sentiment analysis of Modern Standard Arabic and the Algerian dialect. A large manually annotated dataset was collected, mainly from Facebook and other Arabic sources. Two main methods were applied: Support Vector Machines (SVM) and Long Short-Term Memory networks (LSTM), using TF-IDF and Word2Vec for word representation. Experiments conducted on a balanced dataset of nearly 50,000 comments showed that SVM outperformed LSTM, achieving an accuracy of 86%, confirming the effectiveness of supervised methods combined with annotated resources for sentiment classification in under-resourced dialects [47].

The study by Benmounah et al. (2023) introduces a deep learning approach for sentiment analysis of Algerian dialect texts using Transformer-based models. The researchers created a custom dataset of 45,000 manually annotated YouTube comments, labeled as positive, neutral, or negative. Advanced preprocessing techniques were applied, including normalization, emoji handling, and character conversion. The model was trained using an Arabic BERT variant and achieved a notable F1-score of 78.38% and an accuracy of 81.74% on the test set. The results confirm the effectiveness of using BERT for processing dialectal Arabic and highlight the value of tailored preprocessing and dataset construction for improving sentiment classification performance in under-resourced dialects [48].

the study by Guellil et al. (2020) explores a novel zero-effort approach to sentiment analysis of the Algerian Arabic dialect by leveraging existing resources from related dialects. Instead of building new datasets, the researchers utilized manually annotated sentiment corpora in Tunisian (TSAC) and Moroccan (ElecMorocco2016) dialects, along with a large unannotated Maghrebi corpus (NorthAfrica). Several machine learning and deep learning models were evaluated, including MLP, LSTM, and biLSTM. The best F1-score of 83% was achieved using MLP with the Tunisian corpus,

and LSTM with a combination of Tunisian and Moroccan corpora. This approach outperformed previous models by 15%, demonstrating the feasibility of cross-dialect resource reuse for sentiment classification in under-resourced Arabic dialects such as Algerian [49]

The study by Guellil et al. (2019) focuses on sentiment classification of Algerian medical forum posts written in a mix of Algerian Arabic dialect and French. The researchers constructed a domain-specific corpus by collecting patient reviews about medications from Algerian online medical forums. After applying standard preprocessing steps such as punctuation removal, lowercasing, stopword elimination, and lemmatization, they used TF-IDF for feature extraction. Multiple machine learning classifiers, including Naive Bayes, SVM, and Random Forest, were evaluated. The experiments, conducted using 10-fold cross-validation, showed that the SVM model combined with TF-IDF features achieved the best performance, with an accuracy of 87%. This work highlights the challenges of processing code-switched text and emphasizes the need for tailored NLP solutions in the Algerian context [50].

Table 2.2: Case of Arabic Algerian Dialect (Latin or Arabic script)

WORK	**Preprocessing**	**feature extraction**	**classification**	**Dataset statistics**	**Evaluation Protocol**	**Evaluation Results**
[46]	- Text cleaning - Arabic-to-Latin transcription - Grouping (phonetic using Soundex + similarity using Levenshtein distance via FuzzyWuzzy)	**- Bag of Words** (Term Frequency) **- Word2Vec models:** CBOW and Skip-Gram	**- Machine Learning (ML) :** • SVM • Naive Bayes (BNB, MNB, GNB) **- Deep Learning (DL) :** • CNN • RNN	11,760 comments (Facebook, YouTube, Twitter) **Versions tested:** • Raw/original dataset • Transcribed (Latin characters) • Enhanced with Word2Vec	90% training 10% testing	The best result was an 83.70% accuracy using the **SVM** algorithm on Latin-transcribed data enhanced with CBOW from Word2Vec.
[47]	Text cleaning + Normalisation	- TF-IDF (for SVM) - Word2Vec CBOW (for LSTM)	•SVM (Support Vector Machine) • LSTM (Long Short-Term Memory)	Comments in Algerian dialect were collected from multiple sources, including: - Facebook (Algerian pages: ElBilad, Hafid Derradji, etc. - Multi-domain dataset by Elsahar et al. (hotels, restaurants, movies, etc.)	85% training 15% testing	**SVM** is better than LSTM in this study (86% Accuracy)
[48]	Text cleaning + Normalisation	Tokenization with **WordPiece tokenizer** (linked to the BERT model)	**- Comparison of traditional models such as :** LSTM and Bi-LSTM with BERT. **- Testing different versions of BERT** (BERT Arabic Base, BERT Arabic Large).	Personalized corpus of 45,000 annotated Algerian YouTube comments	70% training 20% testing 10 % validation	The best model: **BERT Arabic Large**, (Accuracy 81.74%.)

[49]	Text cleaning + Normalisation	- Bag of Words (BOW) - Word2Vec - Doc2Vec - FastTex	**- 4 machine models were tested:** Naïve Bayes (NB) - Logistic Regression (LR) - Random Forest (RF) - Support Vector Machines (SVC) **4 deep models were tested:** - Multi-Layer Perceptron (MLP) - Convolutional Neural Networks (CNN) - Long Short-Term Memory (LSTM) - Bidirectional LSTM (Bi-LSTM)	The study is based on 3 main datasets: . TSAC corpus (Tunisian dialect) . ElecMorocco2016 corpus (Moroccan dialect) . North Africa corpus (Maghrebi dialects)	**For traditional machine learning** 100% of the training . **For Deep Learning Models:** 80% for training 20% for testing	The best model: **LSTM and MLP**, (Accuracy 88.74%.)
[50]	Text cleaning + Normalisation	- TF-IDF (Term Frequency-Inverse Document Frequency)	- Naive Bayes - Support Vector Machine (SVM) - Random Forest	- Corpus from Arabic-speaking medical discussion forums	- Cross-validation	- the best result obtained is with SVM and TF-IDF (87%)

2.7.3 Case of Modern Standard Arabic

In the realm of contemporary Modern Standard Arabic sentiment analysis, [51]. In this work, two distinct corpora were used. The first corpus, OCA, is already available and freely accessible, while the second corpus, SANA, comprises 718 comments collected from three Algerian newspapers and annotated by two Algerian Arabic native speakers, with annotation agreement measured using the kappa coefficient. Sequentially, a preprocessing step was applied, including translating comments in French, English, or Arabizi into MSA, tokenization, stemming, stop word removal, and defining n-grams and word vectors using methods such as TF, TO, TF-IDF, and BTO. For training, they employed SVM with a linear kernel, Naïve Bayes, and k-nearest neighbors (k=9), using 10-fold cross-validation to prevent overfitting. The best accuracy for the OCA corpus was 89% with Naïve Bayes using Tri-gram and TO encoding, while the SANA corpus achieved 75% accuracy with Naïve Bayes using Bigram and TF encoding. A notable finding was that light stemming's benefits varied based on the corpus characteristics [51].

The study by Nabil et al. (2015) introduces LABR, the largest sentiment analysis corpus for Modern Standard Arabic, containing over 63,000 book reviews collected from Goodreads. Each review is rated on a 1–5 star scale. The dataset addresses the lack of Arabic NLP resources and provides standard splits for training, validation, and testing, supporting both sentiment polarity (positive, negative, neutral) and rating classification tasks. A variety of classical classifiers (SVM, Naive Bayes, Logistic Regression, etc.) were evaluated, with SVM and Logistic Regression achieving top performance, surpassing 70% accuracy on sentiment polarity tasks. Additionally, a domain-specific sentiment lexicon was automatically extracted, demonstrating strong effectiveness while drastically reducing feature dimensionality. The study emphasizes LABR's value as a foundational benchmark for future Arabic NLP research [52].

The study by Cherif and Belahbib (2018) presents a supervised machine learning approach for the automatic classification of scientific abstracts in the field of computer science. The authors constructed a custom dataset consisting of abstracts collected from digital libraries such as Scopus and IEEE Xplore, which were manually annotated into predefined categories (e.g., Artificial Intelligence, Databases, Networking). The preprocessing pipeline included PDF-to-text conversion, stop word removal, and lemmatization using Stanford CoreNLP. For feature extraction, Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Dirichlet Allocation (LDA) were employed to represent textual content. Classification was carried out using Support Vector Machines (SVM), Naive Bayes, and Random Forest classifiers. A 10-fold cross-validation protocol was used to evaluate performance. The best results were achieved using SVM with TF-IDF features, reaching an accuracy of 92%. The study highlights

the effectiveness of classical machine learning techniques combined with robust text preprocessing and feature engineering in the domain-specific classification of scientific literature [53].

Table 2.3: Case of Modern Standard Arabic

WORK	**Preprocessing**	**feature extraction**	**classification**	**Dataset statistics**	**Evaluation Protocol**	**Evaluation Results**
[51]	- Text cleaning - Normalisation	- Extraction based on: → Unigrams, Bigrams, Trigrams → Use of weights: TF, TO, TF-IDF, BTO	- SVM (Support Vector Machine) - Naïve Bayes (NB) - k-Nearest Neighbors (KNN)	**- SANA:** Created from three Algerian newspapers (Echorouk, Elkhabar, Ennahar) **- OCA:** Existing corpus of film reviews	10-fold Cross-Validation	**Naïve Bayes** achieved an accuracy of up to 89.80% on the OCA dataset.
[52]	- Text cleaning - Tokenization (Qalsadi)	- Extraction based on: - N-grams (unigrams, bigrams, trigrams) - Tf-Idf (Term Frequency-Inverse Document Frequency)	- SVM (Support Vector Machine) - Logistic Regression - Multinomial and Bernoulli Naïve Bayes - K-Nearest Neighbors (KNN) - Linear Perceptron - Stochastic Gradient Descent (SGD) - Passive-Aggressive Classifier	**LABR** (Large-Scale Arabic Book Reviews)	60% Training 20% Validation 20% Testing	-SVM and Logistic Regression achieved the best performance (75%)
[53]	- Text cleaning - Normalisation of Arabic text	- Bag of Words (BoW)	- Naïve Bayes - Multilayer Perceptron (MLP) - Convolutional Neural Network (CNN)	Corpus of 25,475 Facebook comments (Algerians)	70% training 15% testing 15% validation	CNN: 89.5%, best result

2.8 Conclusion

In short, this chapter has thoroughly looked at, Ai, machine learning as well as deep learning where sentiment analysis has been over-viewed in the context of the Arabic language. We selectively presented a table with the most relevant research papers; each sub-field in the field of sentiment analysis for Arabic was represented with different aspects and perspectives. As for now, the successive part will pass throughout summary of that literature to presenting our design and model for sentiment analysis in Arabic context. In the next chapter, we will focus on methodology, techniques and breakthrough of our approach. And we will explain everything based on foundations of this chapter. Unlike other sentiment analysis models, through painstaking trial and error and evaluation we aim to demonstrate the precision and relevance of our model in analyzing the hidden sentiments embedded in Arabic language data.

Chapter 3: Methodology and Results

3.1 Introduction

Designing robust, maintainable, and scalable systems relies on the use of specific techniques and methodologies. Among the most effective is the modular approach, which involves dividing the system into smaller, manageable components that can be developed, tested, and maintained independently.

In this chapter, we will present the overall architecture of our system, detailing its design principles and modular structure, as well as how these modules interact and integrate to form a cohesive whole. We will also delve into each stage of the architecture, providing an in-depth explanation of its role and functionality.

3.2 Contribution

Generally, for building a sentiment analysis system or any other similar system, we need to divide our work into several steps, which allows us to work with each step separately and fix any issues that may appear in the future in a non-complicated manner. The global architecture of our system is composed of several layers or steps, each layer takes an input and provides an output that is needed in the next layer. Figure 3.1 shows this global architecture as a single entity

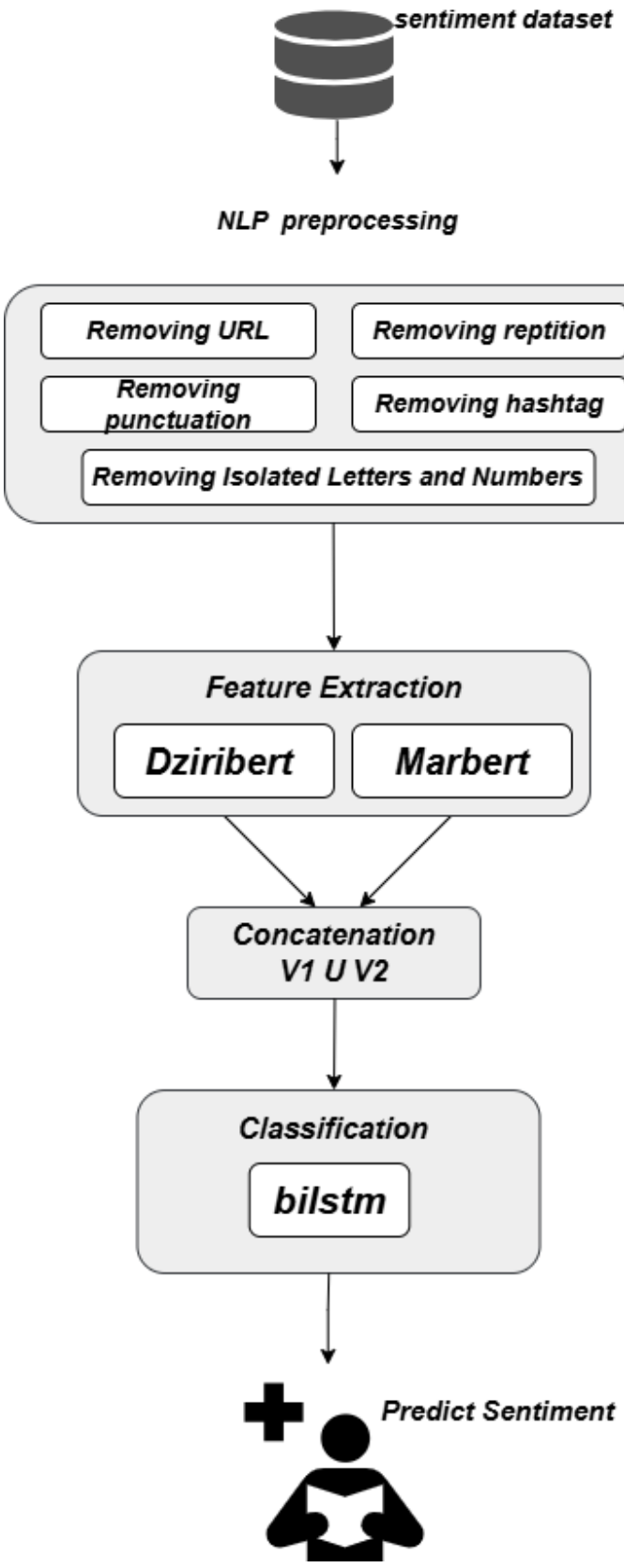


Figure 3.1: Global proposed architecture.

3.2.1 Data overview

In this SA work we will need a set of data represented in a textual form that comes as csv/tsv or simple txt extension as a raw dataset. So we have collected two data sets from different sources and relatively different content, but all of them fall into the field of product reviews of any type, our data is pre-labled according to the sentiment of the review or text as positive or negative or Neutral, and the table bellow represents the collected data with their metadata.

Table 3.1: Datasets collection metadata

Dataset name/label	**Dataset size**	**Total Reviews**	**Class**	**Source file name/type**
Djezzy	709.9 kB	5320	Positives : 567 Neutrals : 2934 Negatives : 1819	Algerian Review.csv
BrandtDz	497.92 kB	6737	Positives : 6737 Neutrals : 6738 Negatives : 3511	BrandtDz.xlsx

Examples from the Dataset one of the most notable characteristics of our dataset is its high linguistic diversity. The collected comments are primarily written in Algerian dialect, but they appear in various forms: some are written using the Arabic script, others in Arabizi (a Latin-script transliteration of Arabic that includes numbers), and many comments include a mix of both forms within the same sentence. Additionally, the French language is frequently used alongside Arabic or Arabizi, reflecting the multilingual reality of Algerian internet users. This variation in writing styles presents a real challenge for natural language processing

- **Djezzy** Some examples of this dataset :

Table 3.2: Dataset "Djezzy "

Comment	**Language Mix**	**Sentiment**
Makanch réseau pendant le weekend	Arabizi + French	Negative
استفيدوا من عروض جازي عايلة الجديدة لي توالم كامل افراد العائلة راهي متوفرة في كامل وكالات جازي	arabe	Neutral
Je suis très satisfait des services de Djezzy	french	Positive
Conx ta3koum dayra ki lhaaaam w say	arabizi	Negative
حابة نرجع بيس تاعي maxi hayla	arabe + Arabizi	Neutral

BrandtDz Some examples of this dataset :

Table 3.3: Dataset "BrandtDz "

Comment	**Language Mix**	**Sentiment**
Dok ngolha ljadati ted3ilk w Rak t3rf fifty fifty	Arabizi	Positive
Awadi mafihach mdr ?? mais bn goood luck ????	Arabizi + anglais	Negative
Bonne nuit wida rbehti hew wajhi	Arabizi + French	Negative
Très bon service, je recommande à 100 %	French	Positive
والله الخدمة تاعهم زيرو، ما نعاودش نثيق فيهم	Arabic	Negative

Dataset structure

- **Djezeyy** The dataset is organized in tabular form with the following columns:

Table 3.4: Structure of the "Djezzyy" dataset

Column Name	Type	Description
name	String	Name or username of the person who wrote the comment.
date	DateTime	The date and time when the comment was posted.
likes	Integer	Number of likes or reactions the comment received
comment	Text	The content of the user's review or message.
sentiment	Categorical	Label annotated with the feeling expressed: **positive**, **negative**, or **neutral**.
source	String	The origin platform or context (Facebook, instgram).

BrandtDz The dataset is organized in tabular form with the following columns:

Table 3.5: Structure of the "BrandtDz" dataset

Column Name	Type	Description
comment	Text	the content of the user's review or message.
sentiment	Categorical	Label annotated with the feeling expressed: **positive**, **negative**, or **neutral**.

3.2.2 Data Preprocessing and preparing :

This step is very important and critical in the sentiment analysis process, as the collected data comes from different sources and exhibits a significant variation in writing styles, especially considering the Algerian context.

Comments are often duplicated and may contain special characters or unwanted words. All cleaning is carried out with the objective of obtaining a corpus with, on the one hand, a minimal vocabulary, and on the other hand, a maximum of informative, meaningful words that convey relevant information regarding polarity.

The details of all preprocessing procedures are provided below:

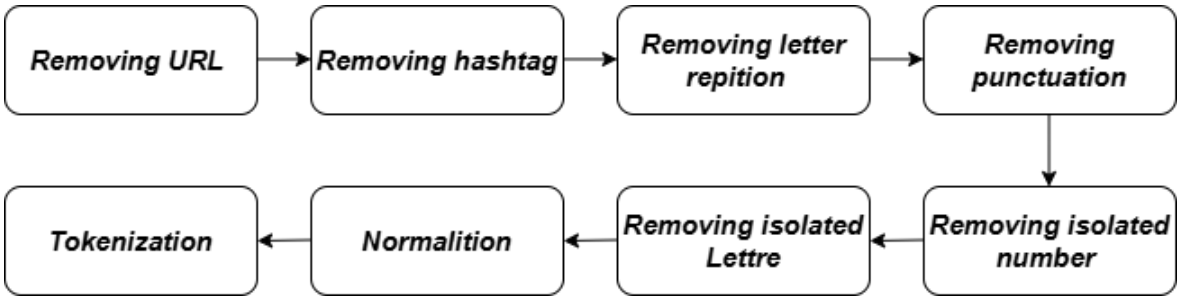


Figure 3.2: Pre-treatment process

The preprocessing phase is identified as the most important step in the sentiment analysis process, as it is during this phase that we prepare the comments in order to build a robust classifier.

As we can observe, on social media, users tend to write in a chaotic manner, and comments may contain spelling mistakes, links, hashtags, GIFs, stickers, special characters, etc., all of which must be removed in order to obtain a minimal, ready-to-use, and unified corpus.

In fact, we keep the steps that are divided between regular and irregular languages . such as :

- **URL Removal:** Any provided URL link generally has no real impact on whether a comment is positive or negative. Therefore, we start by removing URLs from the comment before proceeding with the remaining text.
Input: [chofo had l'article https:// tfahmo bzf swal7]
Output: [chofo had l'article tfahmo bzf swal7]

- **Removing Punctuation:** Punctuation marks, such as "! , . : () ;% - ... " and many others, typically lack meaningful significance in text data. Therefore, removing them is a crucial preprocessing step to reduce noise and simplify the text
Input: [ya weldi hna lmohim mocharka????rba7na f sa7 hata nrb7o hna??]
Output: [ya weldi hna lmohim mocharka rba7na f sa7 hata nrb7o hna]

- **Removing Hashtag :** Hashtags are very representative in sentiment analysis as they can easily shift the polarity of a sentence. We completely removed the hashtag symbol from our corpus and kept only the accompanying word.
Input: [mradh ntouma Djezeyy]
Output: [mradh ntouma]

- **Repetition Removal:** We removed letters repeated more than twice in a word by reducing them to a single instance. The importance of this step lies in reducing the

corpus size. Even though repetition can carry stronger sentiment, in a lexicon-based approach, words written with varying numbers of repeated letters but meaning the same are stored as the same word.
Input: [Waaaaaaaaw ces bien]
Output: [Waw ces bien]

• **Removing Isolated Letters and Numbers:** The goal of this step is similar to removing stop words, except that this part doesn't require a list beforehand; these words don't contribute to the corpus and don't have the meaning to be analyzed, represented, and then calculated and taken into account. And since they are not meaningful, neither are the numbers nor the letters, they will both be removed at this stage.
Input: [b ch7al edha offre]
Output: [ch7al edha offre]

• **Removing diacritics:** Diacritical marks are also eliminated in this step because these signs are not generally used in Arabic texts, and their presence causes ambiguity.
• Replace emoticons with their polarity: However, analyzing emojis presents certain biases, notably that of the subjectivity of the person interpreting the meaning of the comments.
Input: [Rakom ghi Tachéklo 3linà <3 <3]
Output: [Rakom ghi Tachéklo 3linà]

3.2.3 Propostion

As part of this work, a hybrid architecture was adopted that leverages multiple text representations generated by modern language models, namely DziriBERT and MARBERT. These representations were combined using the Late Fusion technique and then fed into a sequential classification model based on BiLSTM. This methodology was chosen based on scientific and technical foundations aimed at improving classification accuracy when dealing with linguistically complex texts, such as Algerian dialect or Arabizi texts converted to Arabic.
DziriBERT is a language model specifically trained on real Algerian data primarily sourced from social media platforms, making it well-suited for understanding the Algerian dialect, including colloquial expressions, informal structures, and local variations of the Arabic language. This type of model excels at capturing local contexts that general models might fail to accurately grasp. Considering that the study's data includes texts converted from Arabizi to Arabic, using a model tailored to the Algerian environment allows for a closer semantic representation of the original intended meaning.
MARBERT is a powerful Arabic language model trained on both Modern Standard

Arabic and various Arabic dialects, especially from social media. It is good at understanding informal and dialectal language, making it a strong complement to DziriBERT. While DziriBERT focuses more on Algerian dialect, MARBERT adds broader dialect coverage and helps capture meaning across different Arabic varieties.

- **Reasons for choosing Late Fusion** Late Fusion is a technique that involves extracting features from each source or model independently, and then combining them at a later stage (typically after the representation phase and before classification). This approach allows each language model to process the data in its own way without interfering with the other models.

Late Fusion was adopted in this work to fuse representations generated by two different language models (DziriBERT and MARBERT) for several academic and technical considerations:

Representational independence: This technique preserves the linguistic specificity of each model. Each text is processed independently by each model, and the results are merged after feature extraction, limiting negative interference between representations.

Contextual and semantic diversity: Late fusion combines the dialectal strength of DziriBERT, which focuses on Algerian Arabic, with the wide dialectal and informal Arabic coverage of MARBERT. This results in richer, more diverse semantic representations.

Simplicity and flexibility: Late fusion is easier to implement than early or intermediate fusion, requiring no changes to model architecture or joint training. It also supports scalability and experimentation with different model combinations.

Model tool variability: Since DziriBERT and MARBERT use different tokenizers and input formats, early fusion is impractical. Late fusion offers a practical and effective solution for combining their outputs.

- **Reasons to choose BiLSTM as the final classifier** The BiLSTM (Bidirectional Long Short-Term Memory) model was chosen as the final classifier for the combined representations generated by DziriBERT and MARBERT, based on several technical and functional properties that make it well-suited for text classification tasks relying on deep contextual embeddings:

Bidirectional Processing: Unlike standard LSTMs that process text in a single direction, BiLSTM processes sequences in both directions—left-to-right and right-to-left. This allows the model to better understand the full context surrounding each word, which is especially important in Arabic due to its flexible word order and rich morphology.

Long-Term Context Retention: BiLSTMs can retain relevant information over long sequences, enabling them to capture relationships between distant words in a

sentence. This helps improve semantic understanding, particularly in longer or more complex Arabic texts.

Suitability for Sequence-to-Classification Tasks: BiLSTM is naturally suited for tasks where an entire sequence is analyzed to produce a single classification output, such as sentiment analysis or text type classification—the focus of this work.

3.2.4 Evaluation metrics

In sentiment analysis, evaluation metrics are essential for assessing how accurately a model can classify text into sentiment categories such as positive, negative, or neutral. Since sentiment datasets are often imbalanced and misclassifications can distort analysis results, it is important to use a variety of metrics for a comprehensive evaluation.

Precision: represents a measure to check the quality of the model's positive predictions It's calculated by looking at the ratio of the correctly predicted positive cases to all the cases the model predicted as positive.

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

Recall: is the proportion of true positive elements identified by the model, out of all the actual positive elements in the dataset. It is calculated by dividing the number of true positives by the sum of true positives and false negatives.

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

Accuracy: is a metric that measures the overall correctness of a model's predictions by comparing the total number of correct predictions to the total number of predictions made. In simpler terms, accuracy tells us how often the model gets its predictions right across all classes or outcomes.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

F1-score: is a metric that combines both precision and recall into a single value, providing a balance between the two. It's calculated by taking the harmonic mean of precision and recall, giving equal weight to both. This makes it a useful measure for evaluating the overall performance of a classification model, especially when the class distribution is uneven or when both false positives and false negatives are important considerations.

$$\text{F1 score} = 2 \cdot \left(\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \right)$$

3.2.5 Results Analysis

We used five state-of-the-art models in Arabic natural language processing, with a particular focus on dialect coverage. The first model, CNN, is a lightweight baseline model suitable for short text classification. It was used in [53] and [45] and achieved the best results.

The article [44] also used the DziriBERT model, a model pre-trained on Algerian dialect data. AraBERT, a popular model in Arabic language processing tasks, was trained on Modern Standard Arabic and a range of its dialects. MARBERT was also used, focusing on Arabic dialects, particularly those spoken in the Gulf region.

Finally, we used the CAMeLBERT model, a dialect-aware, cross-linguistically designed model developed by the CAMeL Lab, which has broad applicability across various Arabic language varieties.

To evaluate the performance of the models, we defined a consistent set of parameters to ensure a fair comparison across all approaches:
Cross-validation: 5-fold
Epochs: 5
Batch size: 32
Learning rate: 1e-4

- **Results with Translation (Arabizi → Arabic)** We followed the state-of-the-art scenario as presented in Articles 1, 2, and 3, where preprocessing steps were applied (such as removing URLs, hashtags, repetitions, isolated numbers, and isolated letters). Then, a translation step from Arabizi to Arabic was performed in order to unify the dataset. Similarly, we applied these steps to the Djezzy and BrandDZ datasets, where the translation was carried out using a phonetic transliteration tool such as Yamli.

djezyy Here are the results after applying the models (dziribert and arabert and marbert and camelbert and CNN) on the data set "djezzy" with translation

BrandDZ Here are the results after applying the models (dziribert and arabert and marbert and camelbert and CNN) on the data set "BrandDZ" with translation

Table 3.6: Results of dataset "djezzy " with Translation

The model	accuracy	precision	recall	macro-f1	f1(negative)	f1(neutral)	F1(positive)
Dziribert	0.7551	0.7118	0.6386	0.6623	0.7331	0.8028	0.4511
Arabert	0.6781	0.7837	0.4742	0.4634	0.6112	0.7557	0.0233
Marbert	0.7510	0.7117	0.6083	0.6310	0.7391	0.7979	0.3559
Camelbert	0.7561	0.6882	0.6519	0.6662	0.7515	0.7983	0.4490
CNN + Bag of word	0.7368	0.7478	0.6017	0.6380	0.6783	0.7967	0.4390

Table 3.7: Results of dataset "brandDZ " with Translation

The model	accuracy	precision	recall	macro-f1	f1(negative)	f1(neutral)	F1(positive)
Dziribert	0.7940	0.6903	0.6432	0.6587	0.3503	0.7572	0.8687
Arabert	0.7716	0.6460	0.6128	0.6238	0.2873	0.7284	0.8557
Marbert	0.7903	0.6753	0.6377	0.6509	0.3351	0.7455	0.8721
Camelbert	0.7940	0.6752	0.6460	0.6570	0.3368	0.7648	0.8695
CNN + bag of word	0.8037	0.6940	0.6491	0.6624	0.3415	0.7692	0.8766

Results Without Translation (Raw Arabizi Text) In this section, we excluded the translation step from Arabizi to Arabic due to the absence of a reliable and validated translation tool. This decision was made to avoid introducing noise or errors into the input, which could negatively affect the performance of the BERT model.

djezyy Here are the results after applying the models (dziribert and arabert and marbert and camelbert and CNN) on the data set "djezzy" Without Translation

Table 3.8: Results of dataset "Djezzy " Without Translation

The model	accuracy	precision	recall	macro-f1	f1(negative)	f1(neutral)	F1(positive)
Dziribert	0.8073	0.7968	0.7862	0.7913	0.8346	0.7637	0.7756
Arabert	0.7353	0.6784	0.6913	0.6840	0.7839	0.5495	0.7184
Marbert	0.7874	0.7811	0.7700	0.7752	0.8131	0.7566	0.7560
Camelbert	0.7850	0.7771	0.7601	0.7681	0.8171	0.7420	0.7454
CNN + bag of word	0.7039	0.6587	0.6645	0.6612	0.7557	0.5622	0.6657

BrandDZ Here are the results after applying the models (dziribert and arabert and marbert and camelbert and CNN) on the data set "BrandDZ" Without Translation

Table 3.9: Results of dataset "brandDZ "Without Translation

The model	accuracy	precision	recall	macro-f1	f1(negative)	f1(neutral)	f1(positive)
Dziribert	0.8103	0.6961	0.6794	0.6867	0.3980	0.7716	0.8905
Arabert	0.7382	0.6472	0.5850	0.6020	0.3212	0.6629	0.8219
Marbert	0.7757	0.6517	0.6405	0.6455	0.3369	0.7439	0.8558
Camelbert	0.7836	0.6621	0.6558	0.6586	0.3574	0.7583	0.8602
CNN + bag of word	0.7054	0.6946	0.6758	0.6689	0.3256	0.7456	0.8187

3.3 Ablation study

In our proposed sentiment analysis framework, we integrate DziriBERT and MARBERT transformer models using a BiLSTM fusion architecture. The core design and training parameters of this architecture are chosen to ensure a balanced trade-off between representation power, computational efficiency, and generalization across sentiment classes (Negative, Neutral, Positive).

We begin by concatenating the [CLS] embeddings from both pretrained transformer models, yielding a combined input vector of 1536 dimensions. This vector is then fed into a BiLSTM layer with a hidden size of 256. The BiLSTM captures contextual dependencies in both forward and backward directions, which is essential in sentiment analysis where word order and long-range dependencies significantly affect sentiment polarity. The bidirectional setting effectively doubles the representational capacity by maintaining both forward and backward hidden states, hence the final feature dimension passed to the classifier is 2 hidden size = 512.

To mitigate overfitting and promote regularization, we apply a dropout rate of 0.3 after the BiLSTM layer. This prevents the model from relying too heavily on specific neurons during training. The final classification is handled by a fully connected layer (fc) projecting the BiLSTM outputs into a num-classes = 3 space, aligning with our sentiment classification goals.

Training-wise, we use a batch size of 32 to balance between stable gradient estimates and GPU memory efficiency. The learning rate is set to 2e-5, a commonly effective setting for fine-tuning transformer-based architectures, especially when used in conjunction with AdamW, a variant of the Adam optimizer that includes decoupled weight decay for better generalization. We further use a linear learning rate scheduler with warm-up, which gradually increases the learning rate at the beginning of training, helping the model to converge more smoothly and avoid sudden weight updates. The model is trained over 5 epochs, a sufficient number given the fine-tuning scenario and the complexity of the model.

This architecture and configuration are specifically tailored for the task of Arabic sentiment classification on dialectal and multilingual social media text, where integrating deep contextual embeddings with sequential modeling (via BiLSTM) enables the system to robustly capture semantic and syntactic cues critical for accurate sentiment prediction.

The following experimental results confirm this analysis, as we present the average accuracy achieved across two different types of data:

Dataset "Djezzy" Evaluation of (DziriBERT + CAMeLbert) and Evaluation of (DziriBERT + MARBERT)

Table 3.10: Fusion with "bilstm "

The model	accuracy	precision	recall	macro-f1	f1(negative)	f1(neutral)	f1(positive)
DziriBERT + MARBERT	0.8336	0.8239	0.8146	0.8190	0.8574	0.7909	0.8087
DziriBERT + CAMeLbert	0.8069	0.8072	0.7967	0.8018	0.8422	0.7744	0.7887

Dataset "BrandDZ" Evaluation of (DziriBERT + CAMeLbert) and Evaluation of (DziriBERT + MARBERT)

Table 3.11: Fusion with bilstm

The model	accuracy	precision	recall	macro-f1	f1(negative)	f1(neutral)	f1(positive)
DziriBERT + MARBERT	0.8258	0.7146	0.6794	0.6926	0.4077	0.7808	0.8892
DziriBERT + CAMeLbert	0.8067	0.7231	0.6827	0.6976	0.4247	0.7787	0.8894

Table 3.12: Fusion with bilstm

Later, to assess the impact of the classification layer on the quality of the results, we reused the same late fusion mechanism but replaced BiLSTM with a CNN layer. Although the performance remained acceptable, it was lower compared to BiLSTM's results. This confirms that late fusion achieves its highest effectiveness when integrated within an architecture capable of capturing the full sequential and contextual structure of the text—something that BiLSTM provides. In contrast, CNN mainly focuses on capturing local positional features, which makes it less capable of interpreting deep contextual interactions between words in sentences containing linguistic and dialectal code-switching.

The following experimental results confirm this analysis, as we present the average accuracy achieved across two different types of data :

- **Dataset "Djezzy"** Evaluation of (DziriBERT + CAMeLbert) and Evaluation of (DziriBERT + MARBERT)

Table 3.13: Fusion with CNN

The model	accuracy	precision	recall	macro-f1	f1(negative)	f1(neutral)	f1(positive)
DziriBERT + MARBERT	0.8053	0.7994	0.7857	0.7922	0.8309	0.7715	0.7743
DziriBERT + CAMeLbert	0.7857	0.7603	0.7784	0.7687	0.8146	0.7360	0.7557

Dataset "BrandDZ" Evaluation of (DziriBERT + CAMeLbert) and Evaluation of (DziriBERT + MARBERT)

Table 3.14: Fusion with CNN

The model	accuracy	precision	recall	macro-f1	f1(negative)	f1(neutral)	f1(positive)
DziriBERT + MARBERT	0.8007	0.6877	0.6820	0.6846	0.4087	0.7643	0.8809
DziriBERT + CAMeLbert	0.7778	0.7485	0.7722	0.7593	0.8075	0.7211	0.7494

BiLSTM Outperforms CNN: The results confirm that BiLSTM significantly outperforms CNN when applied with the late fusion technique, whether combining DziriBERT with AraBERT or with CamelBERT. This superiority is attributed to BiLSTM's ability to capture contextual information in both directions, enabling a deeper understanding of the full word sequence and the contextual relationships within the text.

and finnaly based on these experiments, we conclude that late fusion is an effective strategy for leveraging the complementarity between language models, especially when

each is trained on a different type of data (Modern Standard Arabic vs. dialect). However, the effectiveness of this fusion directly depends on the classification layer used. BiLSTM has demonstrated superior performance in this context due to its ability to capture full dynamic contextual representations, which is crucial for multi-source fused representations. In contrast, CNN remains effective for tasks with limited context, but it is less suitable in cases that require a comprehensive understanding of sentence-level context, as is the case in multi-dialect text classification.

3.4 Conclusion

This chaptre demonstrated the effectiveness of late fusion in enhancing sentiment analysis for Algerian texts written in both Arabic and Arabizi. By combining representations from dialect-aware models like DziriBERT, MARBERT, and applying a BiLSTM classifier, the model achieved strong performance, particularly in handling contextual nuances. BiLSTM outperformed CNN due to its ability to capture sequential dependencies .

General Conclusion

This work addresses the use of artificial intelligence techniques in sentiment analysis for the Algerian dialect. We proposed a new methodology based on merging the representations generated by two Transformer-type language models, DziriBERT and MarBERT, using the Late Fusion technique, then passing this combined representation to a classifier based on a BiLSTM network.

The proposed methodology was evaluated on two Algerian dialect datasets: Djezzy and BrandDZ. The results demonstrated the effectiveness of this approach, achieving an accuracy of 0.83 on the Djezzy dataset and 0.81 on the BrandDZ dataset. These results clearly outperformed the performance of each model individually, with the DziriBERT model achieving an accuracy of 0.75 on the Djezzy dataset and 0.79 on the BrandDZ dataset, while the MarBERT model achieved an accuracy of 0.77 on the Djezzy dataset and 0.79 on the BrandDZ dataset.

These results confirm that merging the linguistic representations generated by different Transformer models can significantly improve sentiment classification performance in the Algerian colloquial context. This approach also enables businesses and organizations to gain a more accurate understanding of users' sentiments and opinions, helping them make quick and informed decisions.

In conclusion, the proposed methodology represents a promising step toward developing effective tools for sentiment analysis in resource-poor Arabic dialects. Future work includes expanding this approach to other Arabic dialects, enriching the dataset with more Algerian-language comments, experimenting with alternative architectures, and optimizing model settings for better performance.

Bibliography

[1] E. Kumar, *Natural language processing*. IK International Pvt Ltd, 2013.

[2] S. Catherine, "Externalisation des audits fournisseurs : Les clés de la réussite," *La Vague*, no. 81, pp. 20–24, avril 2024. [Online]. Available: https://www.a3p.org/art-comprendre-langage-evolution-traitement-automatique-langage/

[3] C. Dong, Y. Li, H. Gong, M. Chen, J. Li, Y. Shen, and M. Yang, "A survey of natural language generation," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1–38, 2022. [Online]. Available: https://arxiv.org/pdf/2112.11739

[4] A. P. Wibawa, F. Kurniawan *et al.*, "Advancements in natural language processing: Implications, challenges, and future directions," *Telematics and Informatics Reports*, vol. 16, p. 100173, 2024.

[5] K. Versteegh, *Arabic language*. Edinburgh University Press, 2014.

[6] S. Harrat, K. Meftouh, M. Abbas, W.-K. Hidouci, and K. Smaili, "An algerian dialect: Study and resources," *International journal of advanced computer science and applications (IJACSA)*, vol. 7, no. 3, pp. 384–396, 2016.

[7] H. Saadane and N. Habash, "A conventional orthography for algerian arabic," in *the Second Workshop on Arabic Natural Language Processing*, 2015, pp. 69–79.

[8] G. Alwakid, T. Osman, and T. Hughes-Roberts, "Challenges in sentiment analysis for arabic social networks," *Procedia Computer Science*, vol. 117, pp. 89–100, 2017.

[9] A. Abbasi and H. Chen, "Applying authorship analysis to extremist-group web forum messages," *IEEE Intelligent Systems*, vol. 20, no. 5, pp. 67–75, 2005.

[10] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[11] B. Saberi and S. Saad, "Sentiment analysis or opinion mining: A review," *Int. J. Adv. Sci. Eng. Inf. Technol*, vol. 7, no. 5, pp. 1660–1666, 2017.

[12] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using bilstm-crf and cnn," *Expert Systems with Applications*, vol. 72, pp. 221–230, 2017.

[13] S. Behdenna, F. Barigou, and G. Belalem, "Document level sentiment analysis: a survey," *EAI endorsed transactions on context-aware systems and applications*, vol. 4, no. 13, 2018.

[14] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *Journal of King Saud University-Computer and Information Sciences*, p. 102048, 2024.

[15] M. Ojeda-Hernández, D. López-Rodríguez, and Á. Mora, "Lexicon-based sentiment analysis in texts using formal concept analysis," *International Journal of Approximate Reasoning*, vol. 155, pp. 104–112, 2023.

[16] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of research and development*, vol. 3, no. 3, pp. 210–229, 1959.

[17] T. O. Ayodele, "Machine learning overview," *New Advances in Machine Learning*, vol. 2, no. 9-18, p. 16, 2010.

[18] L. Hu, J. Chen, J. Vaughan, S. Aramideh, H. Yang, K. Wang, A. Sudjianto, and V. N. Nair, "Supervised machine learning techniques: An overview with applications to banking," *International Statistical Review*, vol. 89, no. 3, pp. 573–604, 2021. [Online]. Available: https://arxiv.org/pdf/2008.04059

[19] G. Patil, V. Galande, V. Kekan, and K. Dange, "Sentiment analysis using support vector machine," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 1, pp. 2607–2612, 2014.

[20] A. Chader, D. Lanasri, L. Hamdad, M. C. E. Belkheir, and W. Hennoune, "Sentiment analysis for arabizi: Application to algerian dialect." in *KDIR*, 2019, pp. 475–482.

[21] P. P. Surya and B. Subbulakshmi, "Sentimental analysis using naive bayes classifier," in *2019 International conference on vision towards emerging trends in communication and networking (ViTECoN)*. IEEE, 2019, pp. 1–5.

[22] İ. Tarımer, A. Çoban, and A. E. Kocaman, "Sentiment analysis on imdb movie comments and twitter data by machine learning and vector space techniques," *arXiv preprint arXiv:1903.11983*, 2019.

[23] E. Elmurngi, "Approches basées sur l'analyse des sentiments et les techniques d'apprentissage supervisé pour des systèmes de réputation robustes dans l'environnement du commerce électronique," Ph.D. dissertation, École de technologie supérieure, 2020.

[24] G. Tripathi and G. Singh, "Sentiment analysis approach based n-gram and knn classifier." *International Journal of Advanced Research in Computer Science*, vol. 9, no. 3, 2018.

[25] G. Aliman, T. F. S. Nivera, J. C. A. Olazo, D. J. P. Ramos, C. D. B. Sanchez, T. M. Amado, and I. Valenzuela, "Sentiment analysis using logistic regression," *Journal of Computational Innovations and Engineering Applications*, vol. 7, no. 1, pp. 35–40, 2022.

[26] Y. Chen, P. Liu, and C. P. Teo, "Regularised text logistic regression: Key word detection and sentiment classification for online reviews," *arXiv preprint arXiv:2009.04591*, 2020.

[27] H. B. Barlow, "Unsupervised learning," *Neural computation*, vol. 1, no. 3, pp. 295–311, 1989.

[28] R. M. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment analysis in arabic tweets," in *2014 5th international conference on information and communication systems (ICICS)*. IEEE, 2014, pp. 1–6.

[29] K. G. Kim, "Book review: Deep learning," *Healthcare informatics research*, vol. 22, no. 4, pp. 351–354, 2016.

[30] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models," *arXiv preprint arXiv:1904.04206*, 2019.

[31] Y. Xie and R. C. Raga Jr, "Convolutional neural networks for sentiment analysis on weibo data: A natural language processing approach," *arXiv preprint arXiv:2307.06540*, 2023.

[32] E. Omara, M. Mousa, and N. Ismail, "Character gated recurrent neural networks for arabic sentiment analysis," *Scientific Reports*, vol. 12, no. 1, p. 9779, 2022.

[33] G. Murthy, S. R. Allu, B. Andhavarapu, M. Bagadi, and M. Belusonti, "Text based sentiment analysis using lstm," *Int. J. Eng. Res. Tech. Res*, vol. 9, no. 05, pp. 299–303, 2020.

[34] M. F. Aslim, G. Firmansyah, B. Tjahjono, H. Akbar, and A. M. Widodo, "Utilization of lstm (long short term memory) based sentiment analysis for stock price

prediction," *Asian Journal of Social and Humanities*, vol. 1, no. 12, pp. 1241–1255, 2023.

[35] V. Chakkarwar, S. Tamane, and A. Thombre, "A review on bert and its implementation in various nlp tasks," in *International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022)*. Atlantis Press, 2023, pp. 112–121.

[36] L. Bharadwaj, "Sentiment analysis in online product reviews: mining customer opinions for sentiment classification," *Int J Multidiscip Res*, vol. 5, no. 5, 2023.

[37] C. N. Chandra, S. Gupta, and R. Pahade, "Sentiment analysis and its challenges," *International Journal of Engineering Research & Technology*, vol. 4, no. 03, pp. 968–970, 2015.

[38] A. A. Q. Aqlan, B. Manjula, and R. Lakshman Naik, "A study of sentiment analysis: Concepts, techniques, and challenges," in *Proceedings of International Conference on Computational Intelligence and Data Engineering: Proceedings of ICCIDE 2018*. Springer, 2019, pp. 147–162.

[39] D. G. Maynard and M. A. Greenwood, "Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis," in *Lrec 2014 proceedings*. ELRA, 2014.

[40] O. Bellar, A. Baina, and M. Ballafkih, "Sentiment analysis: predicting product reviews for e-commerce recommendations using deep learning and transformers," *Mathematics*, vol. 12, no. 15, p. 2403, 2024.

[41] R. N. Ganji, C. Dadkhah, and N. Tohidi, "Improving sentiment classification for hotel recommender system through deep learning and data balancing," *Computación y Sistemas*, vol. 27, no. 3, pp. 811–825, 2023.

[42] C. N. Dang, M. N. Moreno-García, and F. D. l. Prieta, "An approach to integrating sentiment analysis into recommender systems," *Sensors*, vol. 21, no. 16, p. 5666, 2021.

[43] I. Karabila, N. Darraz, A. El-Ansari, N. Alami, and M. El Mallahi, "Enhancing collaborative filtering-based recommender system using sentiment analysis," *Future Internet*, vol. 15, no. 7, p. 235, 2023.

[44] A. Abdedaiem, A. H. Dahou, M. A. Cheragui, and B. Mathiak, "Fassila: A corpus for algerian dialect fake news detection and sentiment analysis," *Procedia Computer Science*, vol. 244, pp. 397–407, 2024.

[45] I. Guellil, A. Adeel, F. Azouaou, F. Benali, A.-E. Hachani, K. Dashtipour, M. Gogate, C. Ieracitano, R. Kashani, and A. Hussain, "A semi-supervised approach for sentiment analysis of arab (ic+ izi) messages: Application to the algerian dialect," *SN Computer Science*, vol. 2, pp. 1–18, 2021.

[46] A. C. Mazari and A. Djeffal, "Sentiment analysis of algerian dialect using machine learning and deep learning with word2vec," *Informatica*, vol. 46, no. 6, 2022.

[47] A. Abdelli, F. Guerrouf, O. Tibermacine, and B. Abdelli, "Sentiment analysis of arabic algerian dialect using a supervised method," in *2019 international conference on intelligent systems and advanced computing sciences (ISACS)*. IEEE, 2019, pp. 1–6.

[48] Z. Benmounah, A. Boulesnane, A. Fadheli, and M. Khial, "Sentiment analysis on algerian dialect with transformers," *Applied Sciences*, vol. 13, no. 20, p. 11157, 2023.

[49] I. Guellil, M. Mendoza, and F. Azouaou, "Arabic dialect sentiment analysis with zero effort.\\case study: Algerian dialect," *Inteligencia Artificial*, vol. 23, no. 65, pp. 124–135, 2020.

[50] I. Guellil, F. Azouaou, F. Benali, A. E. Hachani, and M. Mendoza, "The role of transliteration in the process of arabizi translation/sentiment analysis," in *Recent Advances in NLP: The Case of Arabic Language*. Springer, 2019, pp. 101–128.

[51] H. Rahab, A. Zitouni, and M. Djoudi, "Sana: Sentiment analysis on newspapers comments in algeria," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 7, pp. 899–907, 2021.

[52] M. Nabil, M. Aly, and A. Atiya, "Labr: A large scale arabic sentiment analysis benchmark," *arXiv preprint arXiv:1411.6718*, 2014.

[53] A. Soumeur, M. Mokdadi, A. Guessoum, and A. Daoud, "Sentiment analysis of users on social networks: Overcoming the challenge of the loose usages of the algerian dialect," *Procedia computer science*, vol. 142, pp. 26–37, 2018.