burkina institute of technology
Educating a New Generation of Leaders

Ministry of Higher Education, Research and Innovation

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE AND ENTREPRENEURSHIP

OPTION: PROGRAMMING

DEVELOPMENT OF A SENTIMENT ANALYSIS ALGORITHM FOR A BRAND'S E-REPUTATION: CASE OF COMPANIES IN BURKINA FASO.

Presented and publicly defended by:

OBÈD GUÉSWENDÉ YAMEOGO

Thesis director: **Dr YAYA TRAORE**
Lecturer in Computer Science

Thesis co-director: **Mr HASSIMI COMPAORE**
MASTER 2, TELECEL

THESIS ID:

ACADEMIC YEAR: 2022-2023

Preface

Burkina Institute of Technology which officially opened its doors on October 15th, 2018 in Koudougou is a leading institution of higher learning in Burkina Faso, dedicated to the training of a new generation of leaders who will transform their technical skills into business opportunities, who have strong ethical values and who have the passion to contribute to the development of their country.

At its inception in 2018, Computer science was the only course of study available on campus. Students are trained to master programming languages focused on web, desktop and mobile application development. Artificial Intelligence preliminary courses like data analytic, machine learning and deep learning are also taught.

In 2019, BIT reinforced its vision with the arrival of a new field of study which was electrical engineering option renewable energy. This teaching branch equipped the students concerned with the engineering skills needed to be specialists in renewable energy systems.

In 2021, BIT added mechanical engineering with a focus on resources and mining or agriculture. The goal of this field is to overcome the lack of technicians in Burkina Faso who are capable of creating, innovating and adapting solutions to the country's realities in the field of mechanisation.
Moreover, to the traditional curriculum, BIT has added entrepreneurship and management courses, it regularly organises workshops on different topics related to entrepreneurship that allow students to seize opportunities around them and to learn about how to run a business, so students are entitled to at least one internship per year.

BIT is in the LMD system, which means that the duration of training in all fields of study has been extended to three years, leading to a Bachelor's degree. It is with this in mind that we have undertaken the present production.

Dedication

It is with a heart full of gratitude that I dedicate this work to my dearest parents because beyond the financial resources they have deployed to provide me with a quality education, it is their daily presence and emotional support that has made all the difference. They have been there for me every step of the way, ready to pick me up when I stumbled, to encourage me when I was discouraged, and to celebrate with me every little victory.

Acknowledgements

The achievement of this thesis was possible thanks to the help of several people to whom I would like to express all my gratitude:

First of all, my thesis director, Mr. Compaoré Hassami for his patience, his availability and especially his judicious advice.

My dearest parents, Yameogo Koubi and Ouedraogo Rasmata, my young brother Yameogo Abdias for their encouragement.

A special thanks to Mr. Lebian Wilfried who reinforced my level in programming.

My friends and comrades who gave me their moral and intellectual support throughout my process.

Abstract

In recent years in Burkina Faso the digital transformation has had a significant impact on businesses, through factors such as the rise of social media, increased internet connectivity and the increased use of smartphones etc.
We are now witnessing a shift in consumer buying habits as consumers increasingly turn to the internet to research products, compare prices and read reviews before making a purchase decision. In view of the increased competition in the market, companies are now realizing the importance of their e-reputation as a best way to attract more consumers. And as any successful business cannot afford to ignore consumer reviews posted on its pages, these reviews will allow it to improve its products or services in order to be relevant in the market and attract more consumers.
However, analyzing customers' reviews and classify them into positive, negative and neutral reviews can be tedious if there are enough of them. Considering these difficulties, the present study proposes to develop a sentiment analysis algorithm for a brand's e-reputation for Burkina Faso context.
Sentiment analysis is then the process of determining how to extract experiences and emotions from the provided dataset. Another name for it is opinion mining. The decision-making process for customers and businesses can be significantly changed by applying sentiment analysis to the reviews. When creating a sentiment analyzer, various approaches are used. Some of the processes in the methodology include data collection, data preparation, and algorithm training We will examine many processes and sentiment analysis approaches in this study and choose the appropriated approach for this thesis.

Keywords:

E-reputation, Sentiment Analysis, Deep Learning, opinion, Machine Learning, Accuracy, NLP

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
E-reputation	Electronic reputation
AI	Articificial Intelligence
ML	Machine Learning
NLP	Natural Langage Processing
CEO	Chief Executive Officer
VK	VKontakte (Russian Facebook)
APIs	Application Programming Interfaces
INSD	National Institute of Statistics and Demography
DL	Deep Learning
TF-IDF	Term Frequency-Inverse text Frequency
NER	named entity recognition
SVMs	Support Vector Machines

Part I

Introduction and State of the art

Introduction

In an increasingly connected world, a company's online reputation is essential. Sentiment analysis uncovers information that can't be discovered in any other way Medhat, Hassan, and Korashy, 2014. Big Data which refers to large and complex datasets that are difficult to process and analyze using traditional data processing techniques, reveals your overall brand reputation and does the same for your competitors. Just having that data will improve your brand marketing strategy.

The advent of artificial intelligence has opened up new perspectives in the field of sentiment analysis. Machine learning (ML) and natural language processing (NLP) techniques have made it possible to extract valuable information from vast quantities of online text data. These technological advances offer researchers and businesses the opportunity to better understand the emotions, feelings and opinions expressed by users on digital platforms.Chapter 1: General information about the project In the specific case of companies in Burkina Faso, there is an urgent need to develop tools and methods for assessing their electronic reputation. As the country experiences continued economic growth and increased adoption of digital technologies, it is essential to understand how consumers perceive brands and how this affects their business success.

Therefore, it is within the framework of obtaining our degree of Bachelor in Software Engineering that this study proposes the development of a sentiment analysis algorithm specifically applicable to Burkinabe businesses using advanced artificial intelligence techniques. The aim is to provide decision-makers and marketers with valuable information about their company's e-reputation in order to make informed decisions and develop effective strategies to strengthen their brand image and optimise their online presence.

To achieve this objective, this thesis will proceed in several stages. First, a literature review will be conducted to examine the key concepts related to e-reputation, sentiment analysis and artificial intelligence in this field. Next, a rigorous methodology will be used to collect and pre-process the data required for sentiment analysis. Sentiment analysis algorithms will then be developed and evaluated using real data from companies in Burkina Faso. Finally, the results obtained will be analysed and practical recommendations will be drawn up for improving the accuracy of that algorithm.

I

General information about the project

I.1. Background to the study

The context of this study is the field of brand e-reputation, with a focus on companies in Burkina Faso. With the advent of social media and online platforms, brands are increasingly exposed to consumer opinions and reviews. Representing a brand's image and perception on the Internet, e-reputation plays a vital role in consumer trust and has a significant impact on a company's success.

According to BrightLocal Obiedat et al., 2022, a company specialising in local marketing and online reputation management, 93% of consumers say that online reviews influence their purchasing decisions. Also, 86% of consumers read online reviews before making a purchase. These statistics underscore the growing influence and importance of online feedback and opinions in consumer decision-making. With the increasing penetration of the Internet and the proliferation of online platforms, consumers have access to a wealth of information and user-generated content that can greatly influence their purchasing decisions.

The proliferation of online reviews and ratings has changed the way consumers gather information and rate products or services. Individuals now have the opportunity to seek and consider the experiences and perspectives of other consumers, rather than relying solely on traditional marketing messages or brand claims. Online reviews and ratings provide valuable insights into the quality, performance and overall satisfaction of a particular product or service.

In Burkina Faso, companies are also facing the growing importance of electronic reputation *Burkina Faso Digital Report 2021 : statistiques sur l'utilisation d'Internet et les médias sociaux — lekiosquedigitalduburkina.com* n.d. Consumers in Burkina Faso are increasingly connecting and expressing their opinions on social networks, online forums and rating platforms. Therefore, it is critical for national businesses to understand and monitor their electronic reputation to better manage their brand image, maintain consumer trust and anticipate potential risks.

In this context, the aim of this study was to develop a sentiment analysis algorithm tailored specifically for Burkina Faso companies. The goal is to automate the electronic reputation assessment process using advanced natural language processing and machine learning techniques. The algorithm is designed to extract and analyze the opinions, sentiments and feelings expressed by consumers about Burkina Faso brands on online platforms.

By developing this algorithm, this study aims to help understand and improve the electronic reputation of Burkina Faso companies. The results will enable companies to better understand consumer perceptions, identify problem areas and take appropriate actions to proactively manage their online reputation.

I.2. Problematic

Given the widespread digital transformation and the increasing reliance on internet services by companies worldwide, managing and monitoring online reputation has emerged as a critical concern. The e-reputation of a brand, which represents its image on online platforms, has a profound influence on sales and customer loyalty. Consequently, it is imperative for companies to continuously monitor their e-reputation and promptly address any issues that may arise. However, this task can be challenging and resource-intensive, particularly for large companies with a substantial online presence.
However, sometimes we are dealing with thousands of customer reviews on each of a company's products, and it is almost impossible for the company to read all the reviews and classify them as positive, negative or neutral. The development of an algorithmic solution for sentiment analysis, customized for the needs of companies in the Burkina Faso context, will greatly assist in constant e-reputation monitoring and overcoming the challenges associated with managing online reputation effectively.

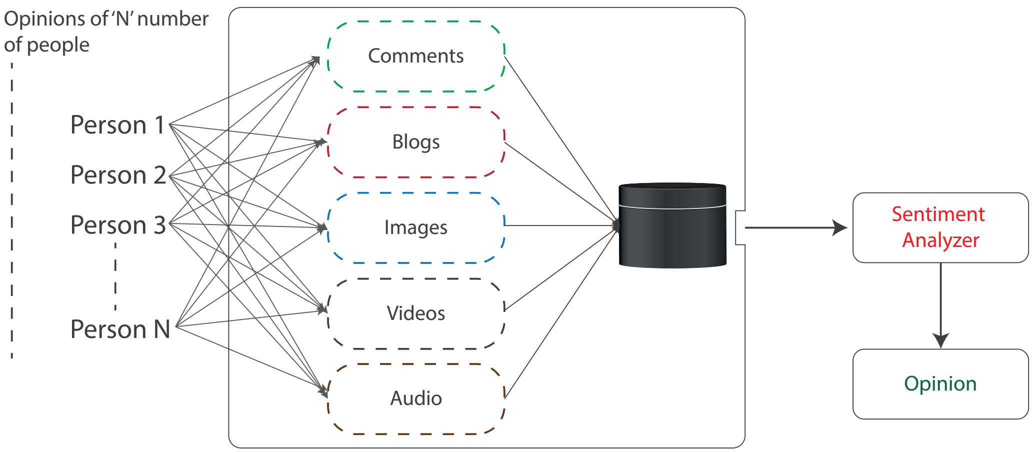

Figure I.1: Data generation

I.3. Objectives

I.3.1. The general objective

The overall objective of this thesis is to develop a sentiment analysis algorithm to help companies in Burkina Faso monitor their e-reputation optimally and effectively to save time and attract more customers as well.

I.3.2. The specific objectives

- **Collecting the data**:

  This objective involves gathering relevant data specific to the Burkina Faso context, such as customer feedback, reviews, social media posts, or any other sources of information that can be used for sentiment analysis.

- **Pre-processing the data**:

  To ensure the quality and reliability of the data, this objective focuses on cleaning, organizing, and preparing the collected data for analysis. It may involve tasks such as removing noise, handling missing data, standardizing formats, and addressing any data quality issues.

- **Developing the model**:

  This objective entails designing and developing a sentiment analysis model tailored to the Burkina Faso context. It includes selecting appropriate algorithms, techniques, and methodologies that can effectively analyze the sentiments expressed in the collected data.

- **Training the model**:

  Once the model is developed, this objective involves training the sentiment analysis model using the pre-processed data. The model learns from the labeled data to understand the patterns and correlations between textual inputs and sentiment labels.

- **Evaluate the model**:

This objective focuses on assessing the performance and effectiveness of the trained sentiment analysis model. we will use some evaluation metrics to measure the model's performance against a validation or test dataset.

- **Suggest ways of improving the model and its application in the context of Burkina Faso**:

  Based on the evaluation results, this objective aims to provide recommendations for enhancing the sentiment analysis model.

This discussed section provides an overview of a research project focused on brand online reputation (e-reputation) in Burkina Faso. The study emphasizes the growing influence of online consumer opinions and reviews on brand perception, highlighting the impact of e-reputation on consumer trust and business success. The surge in online reviews and the role of sentiment analysis are outlined. To carry out this study properly, it would be necessary to study the subject thoroughly and explore the potential predecessors who have proposed solutions along these lines at both international and national level.

II
State of the art

II.1. State of the art

II.1.1. The concept of e-reputation and its importance

While it's interesting to observe the enthusiasm for e-reputation, it's still a little-studied concept, with professionals defining its contours, which is why there's a plethora of expressions to describe it: "Digital awareness," Digital Reputation". ", "online reputation," "online reputation," etc. Electronic reputation is a hybrid concept that explains the myriad of terms used to describe it. But beyond this semantic ambiguity, the definition of electronic reputation also draws on management science , economics, information and communication sciences, etc., which we will consider in more detail in this work . . . as well as previous concepts such as image, identity, and of course reputation, which need to be introduced concepts to understand the problems and risks that "Web 2.0", "Participatory Web" or "Social Networking" bring to e-mail. Company reputation.

According to Fombrun and Shanley Fombrun and Shanley, 1990, an organization's reputation is defined by the public's evaluation of that organization relative to other organizations. In this sense, does a company's reputation represent the public's emotional or emotional response to its name? Therefore, electronic reputation can be defined as the result of perception, interaction, feeling and evaluation of all persons involved in information stimuli (texts, images, videos, etc.) involving a company and disseminated on the Internet. This re-recording of information content gives a company a certain image and ultimately influences consumer behavior.

In the digital age, where the Internet is a key tool for communication and information transmission, the value of e-reputation has grown significantly. Here are a few justifications for why e-reputation is crucial:

- **First impressions**: Nowadays, most people investigate potential interactions with a person, business, or brand online. People may choose not to interact with or purchase goods or services if a company has a bad online reputation.

- **Decision-making process**: An significant role for online reputation in decision-making. When choosing products, services, or brands, consumers frequently rely on online reviews and user opinions. Negative e-reputation might have an impact on purchasing decisions.

- **Professional opportunities**: For people, having a good online reputation can make it easier to get work and pursue professional opportunities. Potential employers frequently review candidate online profiles to gauge their reputation and fit with the company's culture.

- **Crisis management**: In the event of a crisis or other issue, a strong online reputation may be able to lessen the harm. An organization or individual who already has a positive online reputation will be better equipped to handle criticism and deal with issues.

- **Influence**: Social media networks play a crucial role in the development and spread of online reputation. The reputation of a brand or of a person may be significantly impacted by online influencers and users with a large following.

II.1.2. Review of existing studies on sentiment analysis for brand e-reputation

Internationally

On an international level, sentiment analysis is in great demand by companies, which has led to the development of solutions that integrate sentiment analysis algorithms to enable them to effectively monitor their e-reputation, and these tools are already available on the market *15 of The Best Sentiment Analysis Tools — monkeylearn.com* n.d. These tools include :

- **Awario**:

  Awario is a social listening (also called social media monitoring) tool, and sentiment analysis is one of its prominent features.

  The tool crawls all major social media networks, news sites, blogs, forums, and other parts of the Web for mentions of any keyword. Usually, the keyword will be your brand or your competitor's brand, but it can also be your product, industry, CEO, or whatever else you choose.

- **Brandwatch**:

  Brandwatch is another social listening and analytics tool that also performs sentiment analysis. An enterprise-level tool, it covers every source you can possibly think of, including specific social networks such as Tumblr and Goodreads, and local social networks such as VK. If there is a source that you need but can't find on Brandwatch, you can request to add it manually.
- **Talkwalker**:

  Talkwalker is a social listening tool that does much more than just listen. Like Brandwatch, it covers an unlimited number of sources, including social networks, news sites, review sites, blogs, forums, and other parts of the Web. Talkwalker claims to have one of the best sentiment analysis technologies for detecting sarcasm. It also shows customer satisfaction trends, uncovering features of your products that are liked and ones that are disliked.

- **Hootsuite Insight**: Hootsuite is a popular social media management platform. Its main focus is on managing a brand's social activity, but audience analysis is also one of its benefits. Accordingly, Hootsuite Insights analyzes social media networks to reveal overall sentiment regarding your brand and the trends that surround your brand on those networks. You can filter the information by demographic, location, and language.

At national level

In Burkina Faso, a solution that has been developed and is available on the market, enabling companies to manage their electronic reputation by analysing the feelings of their customers, does not yet exist. The general objective of this thesis, as stated, is to fill the existing gap by developing a solution for companies in Burkina Faso to manage their electronic reputation through sentiment analysis of customer feedback. This objective aligns with the need for a localized solution that takes into account the specific language, cultural nuances, and context of Burkina Faso.

By developing this algorithm, we will provide companies in Burkina Faso with the means to monitor and analyze customer sentiments effectively. This can enable businesses to gain insights into customer perceptions, identify areas for improvement, and make informed decisions to enhance their e-reputation.

II.1.3. Gaps in existing research

When considering the gap between internationally and nationally developed algorithms and their applicability to electronic reputation monitoring in the context of Burkina Faso, several factors should be considered:

- **Language and Linguistic Differences** :
  These algorithms are typically trained primarily with data in major language such as English, which can lead to lower accuracy and a lower understanding of feelings expressed in local. And Burkina is mainly a French country so it will important to train algorithms with french data.

- **Cultural context and references** :
  Sentiment analysis algorithms trained on data from international contexts may fail to capture cultural references, idiomatic expressions, or local sentiments specific to Burkina Faso. This can lead to misunderstanding or inaccuracy when analyzing the feelings expressed by the people of Burkina Faso.

- **Social media and platform preferences**:
  Burkinabes are likely to use various social media platforms or online forums that are popular locally. Sentiment analysis algorithms can primarily be trained on data from global platforms, resulting in limited coverage or understanding of sentiment expressed on local platforms.

The transition from traditional reputation management to online platforms is explored, highlighting factors like first impressions, decision-making processes, professional opportunities, crisis management, and social media influence. The review of existing studies presents international and national sentiment analysis tools, revealing a gap in localized solutions. The lack of tailored algorithms for Burkina Faso, considering language, cultural context, and platform preferences, sets the stage for the discussion on the theoretical foundations of sentiment analysis in the following chapter

Part II

Theoretical foundations

III
Methodology

III.1. Data collection

Data sources

For diverse objectives, data can be gathered from a variety of platforms and websites. Here are some typical data sources:

- **Platforms for social media**: Social media data from sites like Facebook, Twitter, and Instagram can be gathered to gain important insights on client sentiment. To compile pertinent postings, comments, and user interactions, leverage APIs or web scraping tools. This information can be processed and analyzed to learn about Burkinabe consumers' attitudes toward particular goods, services, or brands as well as their degrees of satisfaction with them.

- **Local Online Forums and Discussion Boards**: Explore local online forums, discussion boards, and community websites that are popular in Burkina Faso. These platforms can provide valuable insights into customer discussions, opinions, and sentiment regarding various products, services, or industries. Web scraping can be employed to gather data from these sources.

- **online scraping**: Web scraping is the process of directly obtaining data from online sites by examining and interpreting their HTML code. It enables you to collect data from websites without APIs or data feeds. However, it is crucial to make sure that you abide by moral and legal rules and respect websites' terms of service and privacy policies when you scrape data from them.

- **Open datasets platforms**: There are open dataset platforms where you can find a wide range of datasets for various purposes like Kaggle, UCI Machine Learning Repository, Data.gov and more...

- **Surveys and Research Reports**: Look for Burkina Faso-based academic studies, market research studies, and surveys that have already been completed and that contain information on client satisfaction or attitude. These resources can offer insightful information about particular markets, brands, or customer groups.

- **Government Reports and Data**: Investigate government documents, reports, and datasets that may contain statistics on consumer sentiment or satisfaction. Government organizations in Burkina Faso, such the National Institute of Statistics and Demography (INSD), may have reports or data pertinent to customer sentiment in particular industries.

For this study in particular, we are going to opt for the use of an open dataset on international platforms such as Kaggle and Data.world, because they provide more data and, given that this data is in English, we will have to translate the dataset to adapt it to the realities of Burkina Faso, given that the most common language is French.

III.2. Data pre-processing

Cleaning and normalizing data

While underlining that Burkina Faso is a French-speaking country, which implies that in order to obtain fairly convincing results, the argorithm we are going to develop should incorporate data in French. Here are some common pre-processing steps Mishra et al., 2020:

- **Tokenization**: It is a process of splitting or dividing the paragraphs into sentences and sentences into words. The tokenizer function that performs tokenization. There are two types of tokenizers: word tokenizer and sentence tokenizer. Word tokenizer tokenizes the sentence into words while sentence tokenizer tokenizes the paragraph into sentences as shown in FigIII.1.

- **Removing Stop words**: Stop-words are words that are not considered or do not contribute in the analysis process. Rather than storing the stop-words in the datasets,

```
[1]: from nltk.tokenize import sent_tokenize

     sentence = "I love ice cream. I also like steak."
     sent_tokenize(sentence)

[1]: ['I love ice cream.', 'I also like steak.']
```

Figure III.1: Tokenization

it is better to remove the stop-words. Some examples on stopwords:

['au',
'aux',
'avec',
'ce',
'ces',
'dans',
'de',
'des',
'du',
'elle',
'en',
'et',
'eux',
'il',
'ils',
'je',
'la',
'le',
'les',
'leur',
'lui',
]

- **Lowercasing**: One typical normalizing step is to change all text to lowercase. This makes it easier to treat words with the same letters but different cases as being the same. The words "Bonheur", "BONHEUR, "bonheur" for instance, would be changed to "bohneur"

- **Removing Punctuation Marks**: Commas, periods, and exclamation points are frequently omitted unless they serve a specific purpose in the sentence's context. Eliminating punctuation makes the text simpler and quieter.

S.No	Stemming	Lemmatization
1	Stemming is faster because it chops words without knowing the context of the word in given sentences.	Stemming is faster because it chops words without knowing the context of the word in given sentences.
2	It is a rule-based approach.	It is a dictionary-based approach.
3	Accuracy is less	Accuracy is more as compared to Stemming.
4	When we convert any word into root-form then stemming may create the non-existence meaning of a word.	Lemmatization always gives the dictionary meaning word while converting into root-form.
5	Stemming is preferred when the meaning of the word is not important for analysis. Example: Spam Detection	Lemmatization would be recommended when the meaning of the word is important for analysis. Example: Question Answer
6	For Example: “Studies” =¿ “Studi”	For Example: “Studies” =¿ “Study”

Table III.1: Stemming vs Lemmatization

- **Removing Numerical Values**: Numerical values that are not pertinent to the study or work at hand, such as dates or currency symbols, can be omitted from the text or replaced with placeholders.

- **Lemmatization and stemming**: Stemming is the process of finding out morphemes or the root words from the derived word. It does not give the actual morphemes but the closest one.

  Lemmatization is the process of finding out morphemes or the root words from the derived word. The lemmatization process gives the actual root word by comparing the stemmed words in its dictionary and gives the closest actual one.

III.2.1. Extraction of relevant characteristics

Choosing important data from reviews and comments is necessary for the extraction of pertinent aspects. Here are some particular feature extraction methods:

- **Extraction of keywords**: Extracting keywords entails finding the phrases or words that are most commonly used and pertinent to your study. The relevance of terms in a text or corpus can be ascertained by examining the frequency distribution of words or by utilizing methods like TF-IDF (Term Frequency-Inverse text Frequency).

- **Identification of identified entities**: Recognize proper nouns mentioned in the comments, such as company names, places, persons, organizations, etc. These entities can be recognized and extracted using named entity recognition (NER) techniques. This can give important insights into the opinions and specific entities that are being discussed.

- **Identification of identified entities:**:Recognize key expressions or ideas that highlight the primary themes, expressed viewpoints, or particular issues covered in the comments. Techniques like text summarization, topic modeling, or keyword extraction algorithms can be used for this. These techniques can aid in locating the key details and offer a succinct summary of the material.

III.2.2. Sentiment analysis: Concepts and techniques

Sentiment analysis is a branch of artificial intelligence (AI) that aims to determine the opinion, attitude or emotional state of a text, speech or commentary Kaur and Mohana, 2015. It is the classification of subjective content to aggregate users' opinions towards a target entity. It has been successfully used to recommend for commercial use.

The three primary kinds of sentiment analysis techniques are classification-based approaches, lexicon-based approaches, and hybrid approaches.

- **Machine Learning Approach**: Classification-based methods divide texts into various sentiment groups using machine learning algorithms. These methods need training data that has been labeled, and each text sample has been manually marked with the appropriate sentiment (positive, negative, or neutral). The model employs the patterns and features that it has learned from the training data to classify new, unread texts. Naive Bayes, support vector machines (SVMs), logistic regression, random forests, and neural networks are common algorithms for sentiment classification.

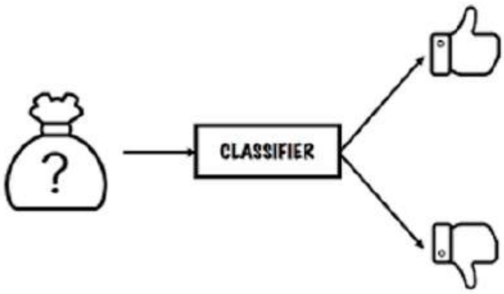

Figure III.2: Machine Learning approach technique

- **Approaches based on lexicons**: Lexicon-based approaches use dictionaries or lexicons that provide predetermined sentiment information for words or phrases. Based on its usage in the lexicon, each word in the text is given a sentiment score. The ratings of individual words are then added up to determine the text's overall mood. To increase accuracy, these lexicons can be strengthened with extra features like negation handling, intensity modifiers, or part-of-speech markers. When tackling sentiment analysis tasks that are particular to a given topic, lexicon-based methods are reasonably quick and efficient.

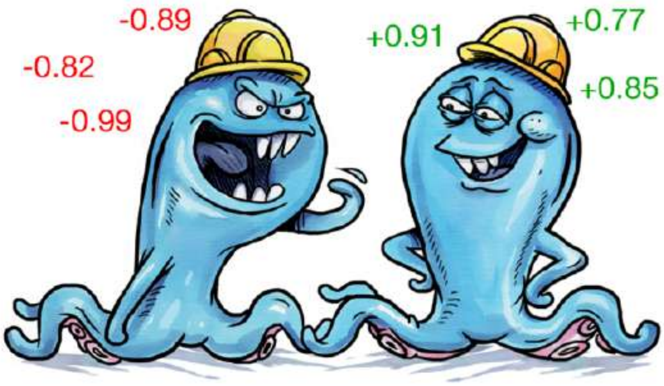

Figure III.3: Lexicon-based technique

- **Hybrid techniques**: Hybrid methods integrate several methodologies to produce sentiment analysis that is more accurate. For instance, a typical hybrid technique uses machine learning algorithms to improve the initial sentiment categorization produced by lexicon-based methods. Combining the two approaches enables taking advantage of their advantages. A different hybrid strategy is to utilize lexicons to extract concepts or entities from the text before using classification algorithms to assess how each concept is perceived separately. Hybrid techniques have the potential to capture complex feelings and enhance performance.

III.3. Sentiment analysis and its areas of application

Sentiment analysis can be used in many different ways. Here are just a few examples:

- **Social media monitoring**: Through sentiment analysis, it is possible to observe the views and attitudes of users in social networks, online forums and blogs. This helps companies understand how their brand, product or service is perceived and adjust their strategy accordingly.
- **Customer service**: Businesses can analyze customer feedback to measure satisfaction and identify potential issues. This enables them to improve customer service and take corrective action quickly.
- **Product rating analysis**: With sentiment analysis, online product ratings can be assessed and the most frequently mentioned positive and negative aspects identified. This helps companies understand consumer preferences and improve their products.
- **Online Reputation Analysis**: Businesses and public figures can use sentiment analysis to monitor their online reputation and assess the impact of positive and negative comments.

III.3.1. Traditional Sentiment Analysis Techniques

These techniques play a significant role in traditional sentiment analysis and have been widely used before the rise of deep learning approaches Pooja Mehta, 02, FEBRUARY 2020. They rely on linguistic knowledge, rules, and dictionaries to extract sentiment from text. While they can provide valuable insights, they may struggle with capturing context, sarcasm, or subtle sentiment expressions, which are areas where more advanced techniques, such as deep learning, have shown improvements.

- **Sentiment Dictionary**: A sentiment dictionary is a list of words that occur with polarity (positive, negative, or neutral). Using these dictionaries, we can assign a polarity to each word in the text and calculate the overall polarity of the text.

- **Rule-based methods**: These methods use linguistic rules to identify feelings and opinions in text. For example, a rule might read: ”If the word ’excellent’ is followed by the word ’product’, then this is a positive review of the product.”

- **Syntactic Analysis**: Syntactic analysis is used to understand the grammatical structure of a sentence. By analyzing the syntax of sentences, we can identify relationships between words and whether words are associated with specific emotions.

III.3.2. Machine learning techniques for sentiment analysis

Machine learning techniques have revolutionized sentiment analysis, enabling higher performance and greater adaptability to different text types. Here are some commonly used machine learning techniques:

- **Supervised classification**: This approach involves using an annotated dataset where each text is associated with a sentiment label (positive, negative, neutral). This data is used to train a machine learning algorithm to predict sentiment labels for new text.

- **Deep Learning** : Deep learning is a branch of machine learning that focuses on teaching artificial neural networks also known as deep neural networks with numerous layers to learn and make predictions or decisions. It tries to use hierarchical levels of abstraction to make it possible for robots to learn and comprehend data representations.

- **Unsupervised learning methods**: These methods do not use previous annotations for training. They rely on techniques such as topic modeling, cluster analysis or anomaly detection to extract sentiment from text.

III.3.3. Comparative study of different approaches and associated advantages/disadvantages

The benefit of using traditional methods for sentiment analysis is that they are generally easy to execute and comprehend. However, because they rely on pre-established rules and lexicons that might not cover all situations, they can be constrained in terms of accuracy and adaptability.

On the other hand, machine learning approaches provide higher precision and adaptability. For specialized jobs, they can be educated on data that is unique to a domain or business, providing more accurate outcomes. They can also pick up on the intricacies and subtleties of the feelings conveyed in words.

Machine learning techniques are favoured when high performance and adaptability to different types of text are needed, whereas classical techniques are frequently utilized for straightforward jobs or when resources are scarce.However, machine learning techniques need a lot of training data, and the learning phase can be time-consuming and resource-intensive. They may also be harder to implement and understand than conventional approaches.

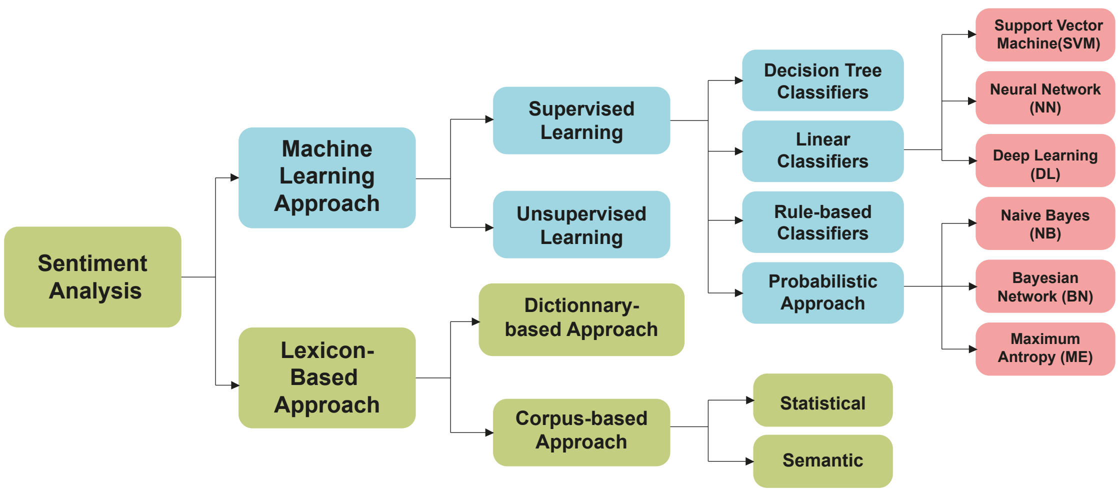

Figure III.4: Sentiment Analysis Techniques

III.4. Development of the sentiment analysis algorithm

The selection of machine learning techniques, the customization of sentiment analysis models for the Burkina Faso environment, and the training and evaluation of the algorithm are all crucial steps in the development of a sentiment analysis algorithm

III.4.1. Choice of machine learning technique

There are a number of machine learning techniques that can be used for sentiment analysis, but we're going to base ourselves more on the machine learning approach, and more specifically on **Deep Learning** Tang, Qin, and Liu, 2015.

Deep learning is a sub-discipline of machine learning that focuses on the use of deep artificial neural networks to solve complex problems Ramadhani and Goo, 2017. It takes its name from the deep structure of the neural networks used. Unlike traditional neural networks, which can be relatively shallow, with only a few layers of neurons, deep neural networks have several hidden layers, enabling them to capture more complex information and features in the data.

Deep neural networks are built by connecting many artificial neurons in successive layers. Each neuron receives weighted inputs, combines them using a non-linear activation function and passes the output to the next layer. The weights and biases of the neurons are adjusted using optimisation algorithms such as backpropagation to minimise the error between the network's predictions and the expected values. It has grown in popularity in recent years due to its ability to process large amounts of data, extract complex features and perform well in a wide range of tasks, including computer vision, natural language processing, speech recognition and more.
As we intend to build a massive data classification model. Deep learning would be the best choice because it has a number of related advantages:

- **Capturing complex relationships**: Deep learning methods, like deep neural networks, are capable of capturing intricate, non-linear correlations between words and emotions. They are able to comprehend the intricacies and subtleties of language by learning hierarchical representations of text data.

- **Great performance**: In many sentiment analysis tasks, deep learning has shown great performance. Deep learning models are able to generalize on new examples and can learn from a lot of data. This makes it possible to make accurate and trustworthy assumptions about the emotions portrayed in texts.

- **Adaptability**: Deep learning models are adaptable to certain tasks and domains. To learn the language and cultural quirks of the area, they can be trained using data that is relevant to the Burkina Faso environment. Due to its versatility, sentiment analysis in Burkina Faso can produce more pertinent and precise results.

- **Relevant word representations**: Deep learning models are able to pick up on rich, distributed word representations, also known as embeddings, that capture the semantic connections between words. These embeddings can capture similarities and contrasts between keywords, representing words in a meaningful way and enhancing sentiment analysis performance.

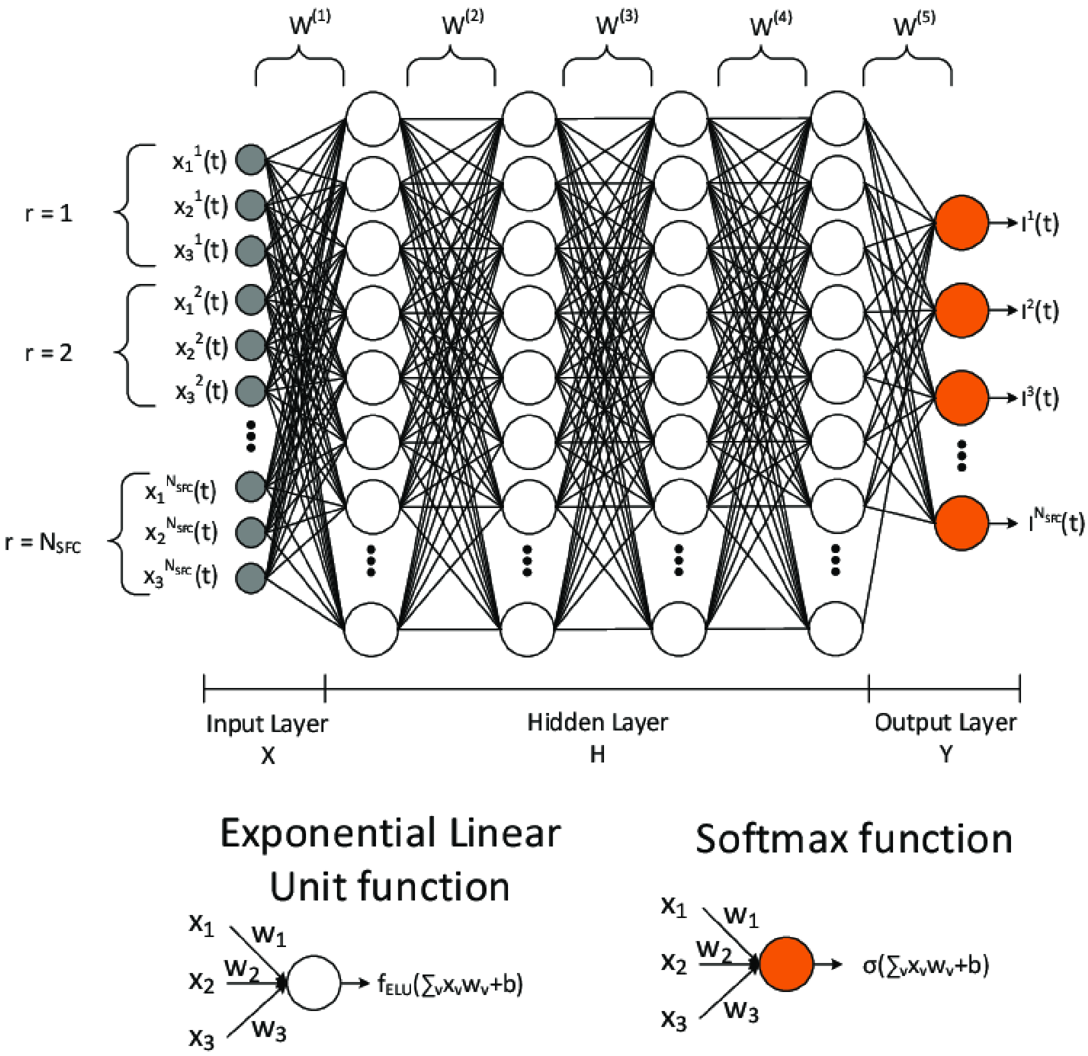

Figure III.5: Deep Neural Network Architecture

III.4.2. Adapting sentiment analysis models to the context of Burkina Faso

The majority of time, data from various contexts and cultures is used to build pre-trained sentiment analysis models. For these models to function best in the Burkina Faso environment, modifications may be necessary. This can entail adding training data that is peculiar to Burkina Faso or modifying the models to account for linguistic and cultural quirks.

III.4.3. Training and evaluation of the algorithm

Once the techniques and models have been selected, it is necessary to gather a set of annotated training data specific to the Burkinabe context. This data can be collected from sources such as social media, online forums or user comments. The algorithm is then trained on this data to learn to predict the sentiment labels associated with the texts.

Evaluating the algorithm is an essential step in measuring its performance. Evaluation metrics such as precision, recall and F-measure can be used to assess the algorithm's ability to accurately predict sentiment in texts from the Burkinabe context. If performance is unsatisfactory, it may be necessary to readjust the models, add more training data or revise the specific features of the Burkinabe context taken into account.

III.4.4. Deployment

Integrate the sentiment analysis algorithm into an application or service. Make sure the model works properly with new data.

This section has enabled us to draw up a roadmap for the development of this algorithm, which will be of great benefit to local businesses. Before we get into the development of the solution, let's take a moment to study exactly why this solution must imperatively meet the requirements of this country... why?

IV

General information on e-reputation

IV.1. Strengths and weaknesses of the e-reputation of Burkina Faso companies

IV.1.1. Strengths

- **Rising online presence**: Businesses in Burkina Faso have a rising online presence, which gives them more visibility and the chance to connect with more people.
- **Social media usage**: Burkina Faso enterprises are increasingly using social media platforms. A company's online reputation can be improved by using social media effectively.
- **Community involvement**: Businesses that actively participate in their online community can increase their credibility and build trust, enhancing their e-reputation.

IV.1.2. Weaknesses

- **Weak usage of social media**: Despite a rising presence, several Burkinabe businesses are still unsure of how social media might help them improve their online reputation.
- **Absence of an online strategy**: Exposure to negative feedback: Burkinabè businesses may encounter bad feedback online. If they are not adequately equipped to handle it, this may harm their online reputation.
- **Exposure to negative feedback**: Burkinabè businesses may encounter bad feedback online. If they are not adequately equipped to handle it, this may harm their online reputation.

IV.2. E-reputation affecting factors

IV.2.1. Analysis of particular factors influencing online e-reputation

- **Product and service quality** : A company's internet reputation may be impacted by how well-received its goods and services are. Positive feedback and a stronger online reputation might result from offering high-quality goods and pleasant customer service.

- **Customer experience**: Online reviews of customers' good or bad experiences with a company can directly affect that company's e-reputation.

- **Transparency in communication**: Honest, up-front dialogue with customers online can boost a business's reputation.

- **Response time**: A company's online reputation may be impacted by how quickly and effectively it reacts to consumer complaints.

IV.2.2. Cultural, economic and social factors linked to e-reputation in Burkina Faso

- **Langage**: How Burkina Faso businesses interact with their audience and shape their online reputation can be affected by the use of French and regional languages on digital platforms.

- **Internet access**: Limited internet access in some regions can restrict the online visibility of Burkinabe businesses, affecting their e-reputation.

- **Trust and reputation offline**: A company's reputation offline can affect its reputation online. This online credibility is advantageous for businesses who already have a strong offline reputation.

IV.3. Recommendations for improving e-reputation

Online communication strategies

- **Develop a social media strategy**: Burkina Faso businesses should develop a clear social media strategy, identifying the most relevant platforms for their target audience and creating engaging, quality content.
- **Monitoring and responding to comments**: It is essential to actively monitor online comments and respond appropriately and promptly, offering resolution to reported issues and thanking positive feedback.

Dealing with negative comments

Respond constructively: When confronted with critical remarks, it's crucial to reply positively by acknowledging the user's concerns and outlining potential solutions or requesting a private dialogue.

Enhancing the brand's reputation:

- **Community involvement**: Actively taking part in online discussions and interacting with the community can improve a company's online reputation and brand image.

- **Quality content**: Producing and disseminating useful, high-quality content can help to boost brand recognition and establish a business as an authority in its industry.

Part III

Implementation of the algorithm

V
Implementation of the algorithm

V.1. Development of the algorithm

V.1.1. Working environment study and choices

To develop the solution, we need to set up a suitable work environment. The requirements for this project are as follows:
A programming language that will enable us to develop our solution. The most popular of these languages are:

Python :
Python is by far one of the most popular languages for developing deep learning solutions, including sentiment analysis. It has many powerful libraries and frameworks such as TensorFlow, Keras, PyTorch and scikit-learn, which simplify the creation, training and deployment of deep learning models.

R :
R is a popular programming language for statistical analysis and data processing. While it may not be as commonly used as Python for deep learning, it has libraries such as Keras and TensorFlow that allow you to develop sentiment analysis solutions.

JS :
JavaScript can be used to develop deep learning solutions, particularly for web applications and interactive user interfaces. TensorFlow.js is a library that lets you run deep learning models directly in a web browser.

And more ...

A machine with sufficient processing capacity. For sentiment analysis, it is important to have a high-capacity machine, especially when working with large datasets or sophisticated machine learning models. A technology (library) that we can use for sentiment analysis using deep learning. As technology adapted to our problem we have a wide range.
There are several technologies and libraries we can use for sentiment analysis using deep learning. Here are some of the technologies commonly used in this field:

Pytorch :
PyTorch is another popular library for deep learning, developed by Facebook. It offers a dynamic approach to computation, which can be advantageous for experimentation and rapid model development.

Keras :
Keras is a popular library for building neural networks in Python. It provides a user-friendly interface for creating and training deep learning models, including those used for sentiment analysis.

And more ...

And obviously a large dataset that we will use to train and test the algorithm.

Tensorflow :
TensorFlow is an open-source library developed by Google for numerical computation and machine learning. You can use it to build, train and deploy deep learning models, including sentiment analysis.

For our part, having made our choices and taking into account the resources available to us, our working environment will be made up of the following elements:

Movie Reviews ....Dataset

For our model, we will use a large dataset of film reviews of 50,000 lines. You can obtain the dataset by following this link `https://drive.google.com/drive/folders/11qLBZfYdC7LLyebnpoAERZhf6-4jFKyK`

HP Pavillon, Processor: CoreI3
Memory(RAM): 16GB
Storage : SSD

For this project we will be using two code editors, Google Colab and Visual Studio code, which are IDEs among many others, such as Jupyter Notebook, Pycharm etc. Google colab for data pre-processing and model training and Visual Studio code for deploying the algorithm in a web application.

Keras because it offers a modular approach to model building and really easy to use

Python as programming because it is the most popular languages for developing deep learning solutions

Google Colab (short for Google Colaboratory) is an online machine learning platform that lets you write and run Python code in a Jupyter Notebook environment directly in the browser. Colab is often used for machine learning, data science and other programming tasks. Here's how to use Google Colab to develop a sentiment analysis algorithm

Visual Studio Code offers a flexible and powerful development environment for working on sentiment analysis projects, with extensions for Python and data science that make it easy to write, debug and visualise your code.

V.1.2. Presentation of source code extracts

Importation of the dataset

To import the dataset we need first to install the prerequisite dependency which is pandas

```
import pandas as pd
```

After importing pandas as a prerequisite dependency for loading our dataset, we need to connect our google drive where the project folder is located to our notebook.

```
from google.colab import drive
drive.mount("/content/drive/")

# And then migrate to our project folder with the command %cd
%cd /content/drive/My Drive/Project/SAproject
```

Stemming and Lemmatization code extracts

```
# Create stemming and lemmatization objects
stemmer = SnowballStemmer(nltk_language)
lemmatizer = WordNetLemmatizer()

def apply_stem(text):
    words = word_tokenize(text)
    # Apply stemming
    stemmed_words = [stemmer.stem(word) for word in words]
    # Join the words back into a single string
    stemmed_text = ' '.join(stemmed_words)
    return stemmed_text

def apply_lemma(text):
    words = word_tokenize(text)
    # Remove stopwords
    words = [word for word in words if word.lower() not in stop_words]
    # Apply lemmatization
```

```
    lemmatized_words = [lemmatizer.lemmatize(word) for word in words]
    # Join the words back into a single string
    lemmatized_text = ' '.join(lemmatized_words)
    return lemmatized_text
```

saving the model and customised trained vectorizer code extract

```
import joblib

# Save the trained vectorizer and model
joblib.dump(vec, 'fitted_vectorizer.pkl')
model.save("model.h5")
```

Saving the model and customised trained vectorizer code extract

The code below allows us to train the model, predict labels for the test data, and evaluate performance using the accuracy and classification ratio. This gives us an idea of the quality of our model for sentiment analysis.

```
model.compile(loss='categorical_crossentropy', optimizer='rmsprop',
metrics=['accuracy'])
label_encoder.fit(['negative', 'positive'])

ytrain_encoded = label_encoder.transform(ytrain)
ytest_encoded = label_encoder.transform(ytest)

ytrain_onehot = np.eye(2)[ytrain_encoded]
ytest_onehot = np.eye(2)[ytest_encoded]

model.fit(Xtrain, ytrain_onehot, batch_size=32, epochs=10, validation_data=
(Xtest, ytest_onehot))
# Predictions and Evaluation
predictions_proba = model.predict(Xtest)
predictions = np.argmax(predictions_proba, axis=1)

# Transform predicted labels back to original
predicted_labels = label_encoder.inverse_transform(predictions)

# Print the Classification Report and Accuracy
print("Classification Report:")
print(classification_report(ytest, predicted_labels, zero_division=1))
print("Accuracy: {:.2f}%".format(accuracy_score(ytest, predicted_labels)
* 100))
```

V.1.3. Deployment of the algorithm

Once we have built and trained our deep learning model, we need to deploy it, i.e. integrate it into a system so that the solution is portable and can be delivered to a customer. There are several frameworks and tools that enable us to do this. Among these tools we have.

Although Django is primarily known as a complete web development framework, it can also be used to deploy templates as web applications. It offers a wide variety of features, which can be useful if your application requires additional functionality outside the template.

Flask is an excellent choice for deploying models. Its flexibility and simplicity make it a good choice for creating REST APIs for our models.

FastAPI is a modern, high-performance framework for creating web APIs in Python. It is designed to be fast to develop and fast to run. It supports automatic data validation, interactive documentation generation and more.

For our part we'll be using flask because as mentionned above flask is an excellent choice for deploying models. Its flexibility and simplicity make it a good choice for creating REST APIs for our models.

V.1.4. Deployment code extract

The following code extract shows the dependencies that will be used to deploy our model. Let's precise here that to deploy a ML model or a Deep Learning (DL) model we have to create a virtual environment which will contain all the dependencies that our project will need to run without the hassle.

```
import base64
from flask import Flask, render_template, request, redirect, url_for
import numpy as np
import pandas as pd
import tensorflow as tf
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk.stem import SnowballStemmer
from nltk.corpus import stopwords
from string import punctuation as punc
import matplotlib.pyplot as plt
import joblib
import io
```

V.1.5. Flask app configuration

Remember that we are using fkask to deploy our model. In Flask before writing your code, you need to create an instance of Flask class because a Flask application is an instance of the Flask class. Everything about the application, such as configuration and URLs, will be registered with this class.

```
app = Flask(__name__, static_url_path='/static', static_folder='static')

# Write your code here
```

And at the end of your code you must add a code of the Flask application to start with debugging enabled. You can access the web application by visiting the URL displayed in the terminal (usually something like http://127.0.0.1:5000/) and interact with it according to your application's routes and logic.

```
if __name__ == '__main__':
app.run(debug=True)
```

V.1.6. Visualization code extract

The following code will return some visualisations thanks to the matplotlib dependency to give an overview of the results of the model's predictions.

```
def create_bar_chart(labels, values):
    colors = [(0, 0.5, 1, 0.5), (1, 0, 0.5, 0.5)]
    plt.figure(figsize=(8, 6))
    plt.bar(labels, values, color=colors)
    plt.title('Distribution des sentiments')
    plt.xlabel('Sentiments')
    plt.ylabel('Comptes')

    # Convertir le graphique en objet Bytes
    img_buffer = io.BytesIO()
    plt.savefig(img_buffer, format='png')
    img_buffer.seek(0)
    return img_buffer

def create_pie_chart(pie_labels, pie_data):
    pie_colors = [(0, 0.5, 1, 0.5), (1, 0, 0.5, 0.5)]
    plt.figure(figsize=(8, 6))
    plt.pie(pie_data, labels=pie_labels, autopct='%1.1f%%', colors=pie_colors)
    plt.title('Répartition des sentiments (Pie Chart)')

    # Convertir le graphique en objet Bytes
```

```
    img_buffer = io.BytesIO()
    plt.savefig(img_buffer, format='png')
    img_buffer.seek(0)
    return img_buffer
```

V.1.7. Some interfaces of the Web App

Here are some interfaces of our web app with the deployed model.

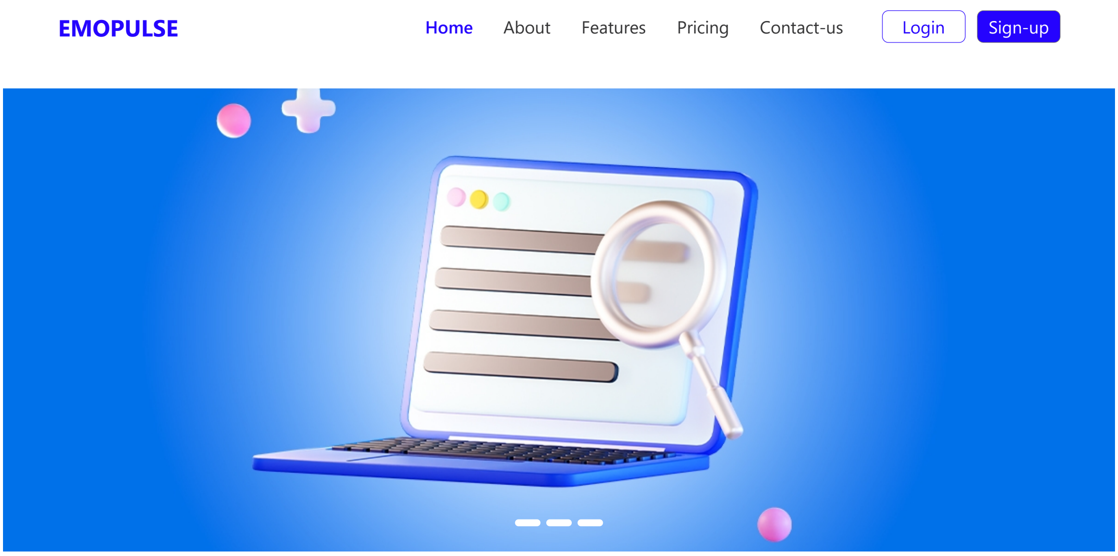

Figure V.1: Home page

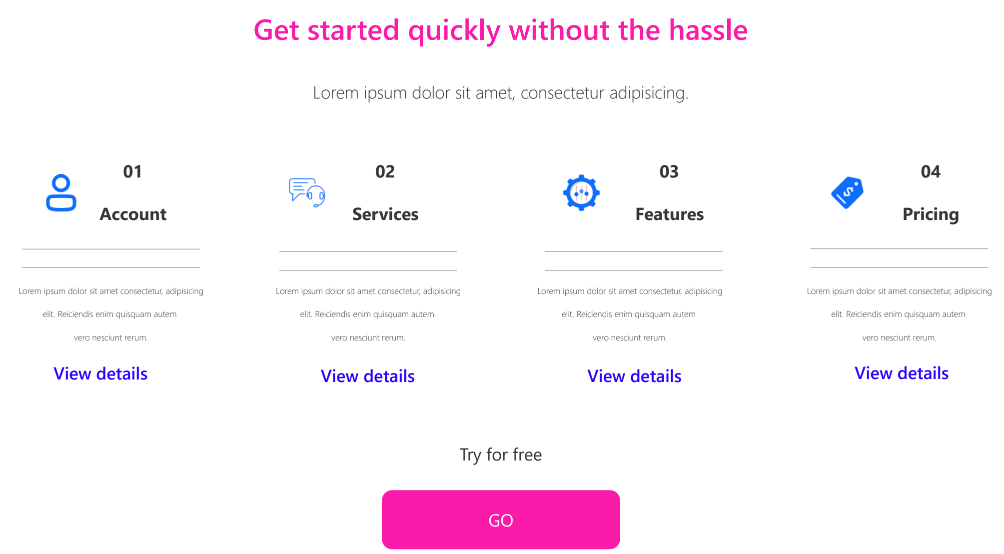

Figure V.2: Guide section

Figure V.3: Information section

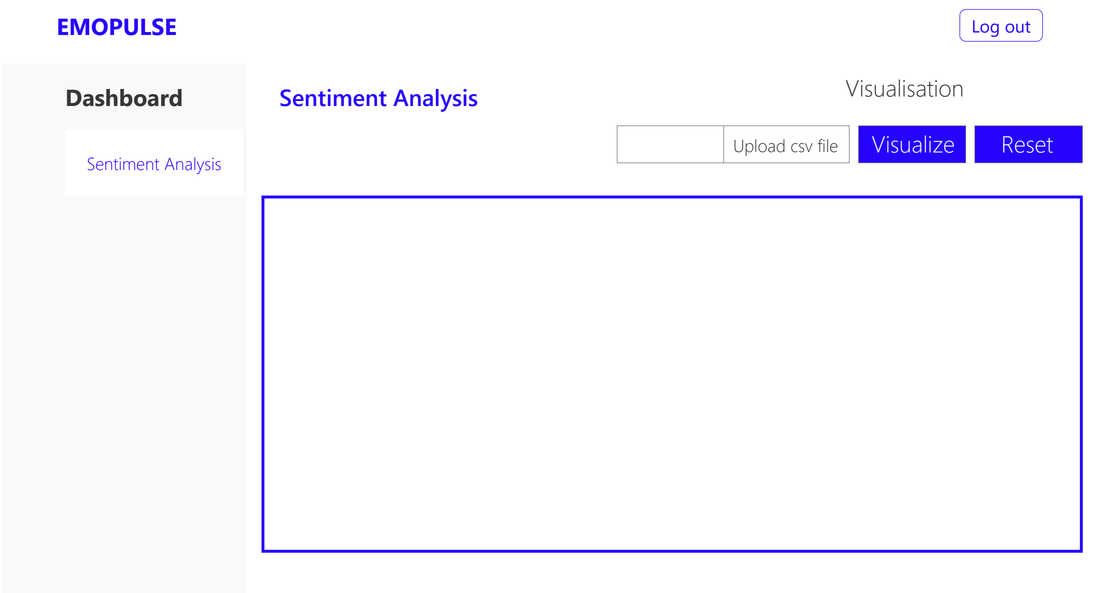

Figure V.4: Visualization section

V.2. Analysis of the results obtained

The results of our classification model indicate an average accuracy of around 72%, which means that your model correctly predicts the class (positive or negative) in around 72% of cases. Here is a more detailed explanation of the metrics in the classification report:

Accuracy (Precision): This is the proportion of correct positive predictions out of all positive predictions. A precision of 72% means that when the model predicts a class as positive, it is correct in around 72% of cases.

Recall: This is the proportion of true positive classes correctly predicted out of all instances of the true positive class. A recall of 72% means that the model correctly identifies around 72% of the true instances of the positive class.

F1-score: This is a measure of precision and recall combined. It is the harmonic mean between precision and recall. An F1-score of 72% indicates a balance between precision and recall.

Support: This is the total number of instances in each class.

Accuracy: This is the total percentage of correct predictions out of the total number of instances. An overall accuracy of 72.05% means that the model correctly predicts approximately 72.05% of all instances.

Macro avg: This is the average of the metrics (precision, recall, F1-score) calculated for each class individually. In this case, the average is around 72%.

Weighted avg: This is the average of the metrics weighted by the number of instances in each class. This average takes into account the imbalance in the distribution of classes. In this case, the average is around 72%.

```
Epoch 1/10
1250/1250 [==============================] - 6s 3ms/step - loss: 0.5789 -
accuracy: 0.6924 - val_loss: 0.5529 - val_accuracy: 0.7206
Epoch 2/10
1250/1250 [==============================] - 3s 3ms/step - loss: 0.5518 -
accuracy: 0.7178 - val_loss: 0.5489 - val_accuracy: 0.7200
Epoch 3/10
1250/1250 [==============================] - 5s 4ms/step - loss: 0.5457 -
accuracy: 0.7230 - val_loss: 0.5459 - val_accuracy: 0.7155
```

```
Epoch 4/10
1250/1250 [==============================] - 3s 3ms/step - loss: 0.5406 -
accuracy: 0.7260 - val_loss: 0.5440 - val_accuracy: 0.7233
Epoch 5/10
1250/1250 [==============================] - 4s 3ms/step - loss: 0.5360 -
accuracy: 0.7285 - val_loss: 0.5481 - val_accuracy: 0.7213
Epoch 6/10
1250/1250 [==============================] - 4s 3ms/step - loss: 0.5314 -
accuracy: 0.7324 - val_loss: 0.5545 - val_accuracy: 0.7235
Epoch 7/10
1250/1250 [==============================] - 5s 4ms/step - loss: 0.5276 -
accuracy: 0.7390 - val_loss: 0.5431 - val_accuracy: 0.7223
Epoch 8/10
1250/1250 [==============================] - 4s 3ms/step - loss: 0.5225 -
accuracy: 0.7398 - val_loss: 0.5543 - val_accuracy: 0.7188
Epoch 9/10
1250/1250 [==============================] - 3s 3ms/step - loss: 0.5184 -
accuracy: 0.7429 - val_loss: 0.5431 - val_accuracy: 0.7232
Epoch 10/10
1250/1250 [==============================] - 4s 3ms/step - loss: 0.5158 -
accuracy: 0.7479 - val_loss: 0.5474 - val_accuracy: 0.7205
313/313 [==============================] - 1s 2ms/step

Classification Report:
          precision    recall  f1-score   support

negative       0.72      0.72      0.72      5000
positive       0.72      0.72      0.72      5000

accuracy                           0.72     10000
macro avg       0.72      0.72      0.72     10000
weighted avg       0.72      0.72      0.72     10000

Accuracy: 72.05%
```

Overall, these results show a balanced performance between precision and recall, with an overall precision of 72%. However, the model still needs to be improved to achieve a high enough accuracy to better predict the Burkinabe context. We should therefore try to adjust the model's hyperparameters, explore other pre-processing techniques and enrich our data to further improve its performance.

V.3. Case studies of specific Burkina Faso companies

This part involves testing the algorithm on data from companies in Burkina Faso. Unfortunately, we will limit ourselves to a single company, Orange Burkina Faso, which is a subsidiary of the telecommunications company Orange, operating in Burkina Faso. It offers a range of telecommunications services, including mobile services, internet and mobile financial services. We have been able to collect customer data for this company which you can access by following this link.

- Distribution of the sentiment
  - 54.7% of the data were classified as positive comments.
  - 45.3% of the predictions were classified as negative comments.
- The most frequent words in the dataset are:
  - ”network”
  - ”connexion”
  - ”good”
  - ”expensive”
- the dataset size is 202

These results reveal that the Orange network is appreciated by most for the fluidity of its Internet connection, but that tariffs remain an important factor influencing customers’ choice, as they find the network rather expensive. This feedback provides a global view of end-users’ feelings and is also beneficial for the company, as it will understand its customers’ needs and can therefore decide to review its tariffs to make the network accessible to all and attract more customers.

Conclusion

In conclusion, this research aimed to shed light on the sentiments expressed in online discussions and comments related to various topics in the context of Burkina Faso. Using advanced natural language processing techniques and sentiment analysis algorithms, we dove into the vast sea of textual data to extract meaningful insights. The results obtained have contributed significantly to our understanding of the sentiments that prevail in online content originating from Burkina Faso.

Thanks to the comprehensive analysis of Orange Burkina customer feedback, we revealed that a significant proportion of the sentiments expressed were positive, representing 54.7% of the data. This indicates a generally optimistic outlook among online contributors, underscoring the potential for positive engagement and growth in the digital space. However, it's worth noting that 45.3% of the predictions lean towards negative feelings, highlighting areas of concern that deserve attention.

Identifying recurring keywords, such as "network," "connection," "good," and "expensive," offers valuable insights into common themes emerging from online discourse. The frequent mention of words such as "expensive" calls for further examination of accessibility issues and economic considerations in the context of Burkina Faso. This is an opportunity for stakeholders to address issues related to cost perception and accessibility.

Although this research makes a significant contribution to understanding feelings in Burkina Faso's online environment, it is essential to recognize some limitations. The study relied primarily on textual data available on public platforms, which may lead to selection bias and exclude the sentiments of less digitally active individuals. Furthermore, the sentiment analysis model did not capture the nuanced cultural and contextual factors unique to Burkina Faso due to the problem of insufficient data. Which also leads us to reflect on ways to enrich the model and manage to put a pipeline which is very crucial to automate and rationalize the complex process of transforming the model into production, reliable and scalable distribution of predictions.

For the future, this study will be better refined, because we will dwell on the collection of data from various sources on the local level to enrich our model and thanks to the pipeline system and the continuous integrations, functionalities will be developed to allow to connect direct business web pages to the pipeline and track real-time analytics. Companies will be able to make better decisions and always stay on the lookout for changes in the behavior of their customers.The results of this thesis contribute to

the evolution of the field of sentiment analysis and provide a valuable resource for understanding the sentiment landscape in Burkina Faso. As the digital landscape continues to evolve, the ability to decipher the sentiment of online discussions is becoming increasingly crucial to making informed decisions and fostering positive interactions.

References

15 of The Best Sentiment Analysis Tools — monkeylearn.com (n.d.). https://monkeylearn.com/blog/sentiment-analysis-tools/. [Accessed 05-09-2023].

Burkina Faso Digital Report 2021 : statistiques sur l'utilisation d'Internet et les médias sociaux — lekiosquedigitalduburkina.com (n.d.). https://lekiosquedigitalduburkina.com/2021/03/14/burkina-faso-digital-report-2021-statistiques-sur-lutilisation-dinternet-et-les-medias-sociaux/. [Accessed 05-09-2023].

Fombrun, C. and M. Shanley (June 1990). "WHAT'S IN A NAME? REPUTATION BUILDING AND CORPORATE STRATEGY." In: *Academy of Management Journal* 33.2, pp. 233–258. DOI: 10.2307/256324. URL: https://doi.org/10.2307/256324.

Kaur, Sukhnandan and Rajni Mohana (2015). "A roadmap of sentiment analysis and its research directions". In: *International Journal of Knowledge and Learning* 10.3, p. 296. DOI: 10.1504/ijkl.2015.073485. URL: https://doi.org/10.1504/ijkl.2015.073485.

Medhat, Walaa, Ahmed Hassan, and Hoda Korashy (Dec. 2014). "Sentiment analysis algorithms and applications: A survey". In: *Ain Shams Engineering Journal* 5.4, pp. 1093–1113. DOI: 10.1016/j.asej.2014.04.011. URL: https://doi.org/10.1016/j.asej.2014.04.011.

Mishra, Puneet et al. (Nov. 2020). "New data preprocessing trends based on ensemble of multiple preprocessing techniques". In: *TrAC Trends in Analytical Chemistry* 132, p. 116045. DOI: 10.1016/j.trac.2020.116045. URL: https://doi.org/10.1016/j.trac.2020.116045.

Obiedat, Ruba et al. (2022). "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution". In: *IEEE Access* 10, pp. 22260–22273. DOI: 10.1109/access.2022.3149482. URL: https://doi.org/10.1109/access.2022.3149482.

Pooja Mehta, Dr.Sharnil Pandya (2, FEBRUARY 2020). "A Review On Sentiment Analysis Methodologies, Practices And Applications". In: *INTERNATIONAL JOURNAL OF SCIENTIFIC TECHNOLOGY RESEARCH, ISSUE* 9. ISSN: 2277-8616.

Ramadhani, Adyan Marendra and Hong Soon Goo (Aug. 2017). "Twitter sentiment analysis using deep learning methods". In: *2017 7th International Annual Engineering Seminar (InAES)*. IEEE. DOI: 10.1109/inaes.2017.8068556. URL: https://doi.org/10.1109/inaes.2017.8068556.

Tang, Duyu, Bing Qin, and Ting Liu (Oct. 2015). "Deep learning for sentiment analysis: successful approaches and future challenges". In: *WIREs Data Mining and Knowledge Discovery* 5.6, pp. 292–303. DOI: 10.1002/widm.1171. URL: https://doi.org/10.1002/widm.1171.

Appendix

Appendix 1:

Project directory structure

```
sa_project_here
├── __pycache__
├── .ipynb_checkpoints
├── env
├── static
│   ├── css
│   ├── img
│   └── portalstyle
├── templates
│   ├── index.html
│   ├── portal.html
│   └── pricing.html
├── app.py
├── fitted_vectorizer.pkl
└── model.h5
```

Appendix 2:

Preprocessing code extract

```
def preprocess_text(text):
    text = text.lower()
    text = text.translate(str.maketrans('', '', punc))
    text = text.strip()
    words = [word for word in text.split() if word.lower() not in stop_words]
    stemmed_words = [stemmer.stem(word) for word in words]
    preprocessed_text = ' '.join(stemmed_words)
    return preprocessed_text
```

Appendix 3:

Function to return dataset summary

```
def create_dataset_summary(dataset, predictions):
    num_entries = len(dataset)
    avg_sentiment = predictions.mean()

    # Calculate the most common words in the dataset
    word_count = {}
    for text in dataset['examen']:
        words = text.split()
        for word in words:
            if word not in word_count:
                word_count[word] = 1
            else:
                word_count[word] += 1
    most_common_words = [word for word, count in sorted(
        word_count.items(), key=lambda item: item[1], reverse=True)[:5]]

    summary = {
```

```
        'num_entries': num_entries,
        'avg_sentiment': avg_sentiment,
        'most_common_words': ", ".join(most_common_words)
    }
    return summary
```

Appendix 4:

Let's precise that we can use Flask that include web APIs. It provides a set of tools and libraries that make it easier to create web applications in Python. Flask allows you to define routes, handle HTTP requests and responses, render HTML templates, and more.

You can use Flask to build web APIs by defining routes that respond to HTTP requests with JSON data or other structured formats instead of rendering HTML templates.

Routes to navigate to web pages

```
@app.route('/', methods=['GET', 'POST'])
def home():
    return render_template('portal.html')

@app.route('/pricing')
def pricing():
    return render_template('pricing.html')

@app.route('/index',methods=['GET', 'POST'])
|
|
|
```