PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY OF MOHAMED BOUDIAF - M'SILA

Faculty of Mathematics and informatics

Department of computer science

N° :……..

DOMAIN : Mathematics and computer science

SPECIALITY: computer science

Field: Networks and TICs

Presented for the graduation of MASTER

By : Chabira Safia

Entitled

Analyzing the sentiment polarity of watched movies based on viewers' comments

Defended on: …./…./2021

Board of Examiners:

Meliouh Amel	Université de M'sila	President
Kadri Said	Université de M'sila	Supervisor
Ghemougui .Abd Essattar	Université de M'sila	Reviewer

Academic year: 2020 / 2021

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY OF MOHAMED BOUDIAF - M'SILA

Faculty of Mathematics and informatics

Department of computer science

N° :……..

DOMAIN : Mathematics and computer science

SPECIALITY: computer science

Field: Networks and TICs

Presented for the graduation of MASTER

By : Chabira Safia

Entitled

Analyzing the sentiment polarity of watched movies based on viewers' comments

Defended on: …./…./2021

Board of Examiners:

Meliouh Amel	Université de M'sila	President
Kadri Said	Université de M'sila	Supervisor
Ghemougui .Abd Essattar	Université de M'sila	Reviewer

Academic year: 2020 / 2021

DEDICATION

.

I dedicate this work:

To the two people who have supported me throughout my studies, sometimes comforting me in difficult moments, and without them

I would not have been able to do this work.

*"My dearest **father** and my dearest **mother**".*

*To my husband **"Djalab Abdelhakim "**. and her family.*

*To my son: **" Mohamed Zineddine "***

*To my brothers **"Abdelmonîam "**, **" Elhassen "**, and sisters **"Zohour "**, **"Halima "**, **"Rayhana "**, **"Kawther "***

my grandfather and my grandmothers.

To all my uncles and their families, my tents and their families.

To all my friends.

.....Safia Chabira

Appreciations

*Thanks to **almighty Allah** -who helped me- to realize this work.*

*I would like to thank my supervisor **Dr.Kadri Said**, Doctor at the University of M'sila, for having proposed, directed, and followed this work.*

My thanks also go to all the teachers and colleagues of studies, who helped me for the realization of this work.

Without forgetting all the staff of the department of computer science department.

Table of contents

Chapter 2: Machine learning and Deep learning

Chapter 3: Environment tools

Chapter 4: Realized work and obtained results

List of figures

List of tables

List of abbreviations

SA: Sentiment Analysis

NLP: Natural Language Processing

DL: Deep Learning

ML: Machine Learning

NPS: Net Promoter Score

VoC: Voice of the customer

POS: Parts of Speech

k-NN:k-Nearest Neighbors

CRF: Conditional Random Field

HMM: Hidden Markov Model

SDC: Single Dimensional Classification

SMO: Sequential Minimal Optimization

AI: Artificial intelligence

SVM: Support Vector Machines

ANN: Artificial Neural Network

SGD: Stochastic Gradient Descent

CNN: Convolutional neural network

RNN: Deep Recurrent Neural Network

GPU: Graphics Processing Unit

CPU: Central Processing Unit

CSV: Comma Separated Values

GENERAL INTRODUCTION

GENERAL INTRODUCTION

"Knowledge of words leads to knowledge of things."

Platon

In recent years, the growth in the amount of data available on the way individuals search for knowledge. In addition to the content made available by the media, users have expanded their role in browsing or in the search for content and began to share their knowledge on the Web experiences, criticism, and opinions on the Web through personal blogs, social networks, and other means.

In this way, the use of the web for data creation and sharing has generated possibilities for analyzing and mining data.

One of the possible applications for the analysis of this shared content is to analyze the sentiment polarity of watched movies based on viewers' comments, such as establishing the opinion that consumers have about watched movies quality. The opinions are important because individuals and organizations are influenced by them when making a decision making In the case of organizations, conducting surveys or focus groups face to face with consumers in order to collect opinions about their products and their competitors' is an activity less and less frequent since nowadays there is an abundance of information and opinions of individuals available on the web. However, before the web existed, there were hardly any computational studies about opinion mining as few texts with user opinions were available.

With the expansion of the participation of users in the web contents exposing their thoughts, new concepts and methodologies are emerged to investigate these contents.

One of the challenges in the area of artificial intelligence, especially in the subarea of natural language processing (NLP), is to computationally treat opinions, feelings and subjectivity in text Sentiment Analysis (SA) is one of the techniques that is increasingly gaining momentum in the treatment of recognition of affectivity inferred from writing.

In this context, dealing with opinions and feelings related to the film is another major challenge in terms of being able to recognize negative and positive ones.

1- Objective

The goal of this project was to build a sentiment classification method, to be applied on watched movies based on viewers' comments.

This method provides a polarity (positive or negative) of the consumer's opinion published content, enabling to offer a sentiment analysis with characterization of aspects of popularity and repercussion of the theme in focus.

2- Manuscript Organization

This document is organized into the following chapters:

Chapter 1: Sentiment Analysis

Chapter 2: Machine learning and Deep learning

Chapter 3: Environment Tools

Chapter 4: The realized work and the obtained results.

CHAPTER 1

SENTIMENT ANALYSIS

1. Introduction

This chapter is an introduction to the project field, in this chapter we talk about sentiment analysis, its definition, importance, and application domain, and their technique.

2. Related work:

This section reviews the work done in the field of text sentiment analysis from three aspects: sentiment analysis methods based on sentiment lexicon, sentiment analysis methods based on machine learning and sentiment analysis methods based on deep learning.

SA aspects	Author	Description
Sentiment Lixicon	Taboada et Al [**1]**	The Semantic Orientation Calculator to extract sentiment from text using dictionaries of words annotated with polarity and strength.
	Jurek et al [**2]**	Proposed a new lexicon-based sentiment analysis algorithm using namely sentiment normalization and evidence-based combination function. In addition to using sentiment terms
	Asghar et al [**3]**	Integrated emotions, modifiers and domain specific terms to analyze sentiment analysis of online user comments
	.Bandhakavi et al [**4]**	Proposed a unigram mixture model (UMM) based DSEL by using labeled and weakly-labeled emotion text to extract effective features for emotion classification
	Dhaoui et al. [**5]**	Used the LIWC2015 lexicon and RText Tools machine learning package to compare the sentiment analysis method based on lexicon and machine learning
	Khoo and Johnkhan [**6]**	Proposed a general sentiment lexicon called WKWSCI Sentiment Lexicon and compares it with the existing sentiment lexicons
	Zhang et al [**7]**	Analyzed text sentiment of Chinese microblog by using extended sentiment lexicons of added degree adverb lexicon, network word lexicon and negative word lexicon
	Keshavarz and Abadeh [**8]**	Used a combination of corpus and lexicons to construct adaptive sentiment vocabulary to improve the sentiment classification accuracy of Weibo.
	Feng et al [**9]**	Constructed a two-layer graph model using emoji and candidate sentiment words, and selected the top words in the model as sentiment words
MACHINE LEARNING	Manek et al [**10]**	Proposed the method of feature extraction based on gini index and classification by support vector machine (SVM).
	Hai et al [**11]**	Proposed a new probabilistic supervised joint emotion model (SJSM), which could not only identify semantic sentiments from the comment data but also infer the overall sentiment of the comment data.
	Singh et al [**12]**	Used naive bayes, J48, BFTree and OneR four machine learning algorithms for text sentiment analysis
	Huang et al [**13]**	Proposed a multi-modal joint sentiment theme model. Based on the introduction of user personality features and sensitive influence factors, the model uses LDA model to analyze the hidden user sentiments and topic types in Weibo text

	Huq et al [**14]**	Used SVM and KNN algorithms to analyze the sentiment of twitter data.
	Long et al [**15]**	Used SVM to classify stock forum posts using additional samples containing prior knowledge.
DEEP LEARNING	Jianqiang et al [**16]**	Used the contextual semantic features and the co-occurrence statistical features of the words in the tweet and the n-gram feature input convolutional neural network to analyze the sentiment polarity
	Hyun et al [**17]**	Proposed a TCNN the model uses the distance relationship between the target word and the surrounding words to learn the influence of surrounding words on the target words. Attention mechanisms arise because each word in a sentence has a different effect on the emotional polarity of the sentence. In order to combine the dominant and recessive features in the sentence
	Ma et al [**18]**	Proposed an extended LSTM called Sentic LSTM. The model unit includes a separate output gate for inserting token level memory and concept level input. Based on the mathematical theory of regression neural network
	Chen et al [**19]**	Proposed the LSTM model for the detailed emotional analysis of Chinese product reviews
	Wen et al [**20]**	Proposed a memristor based MLSTM network hardware design using memristor crossbars.
	Abid et al [**21]**	Proposed a joint structure that combines CNN and RNN. The structure uses the RNN to locate the CNN and uses the global average pool layer to capture long-term dependencies with CNN
	Chen et al [**22]**	Proposed a divide-and-conquer method, which first uses a neural network-based sequence model to classify sentences, and then inputs each set of sentences into a convolutional neural network for sentiment classification.
	Hu et al [**23]**	Performed sentiment analysis of short texts by constructing a keyword vocabulary and combining the LSTM model.
	L. Yang et al **[24]**	Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning.

Tab1e 1.1 A comparative study of different logic classification algorithms

3. Sentiment analysis

Sentiment analysis is a type of data mining. It applies a mix of statistics, natural language processing (NLP), and machine learning to identify and extract subjective information from text file, for instance, a reviewer’s feelings, thoughts, judgments, or assessments about a particular topic, event, or a company and its activities.[25] Opinion mining (with a focus on extraction) or emotional rating are other terms for this sort of study. The words sentiment segmentation and extraction are also used by certain experts. The goal of sentiment analysis, regardless of the label, is the same: to determine a user's or audience's opinion on a target item by evaluating a large volume of text from numerous sources.[26]

In English dictionaries, Sentiment is an attitude toward something; regard; opinion. and refined or tender emotion; manifestation of the higher or more refined feelings. while the analysis is this process as a method of studying the nature of something or of determining its essential features and their relations.[27]

4. Importance of Sentiment analysis

From the definition of sentiment analysis mentioned above:

- We can fix the main goals of its study as follows : To define automatic tools that extract subjective information from natural language text, such as opinions and feelings, to create opinions and feelings, to create structured knowledge that can be used by a decision support system or by a decision-maker.[28]
- Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. Social media monitoring tools like Brandwatch Analytics make that process quicker and easier than ever before.
- Sentiment analysis is critical because it allows firms to immediately grasp the state of the market.

5. The domain of Sentiment analysis

we've listed some of the most popular ways that sentiment analysis is being used in business:

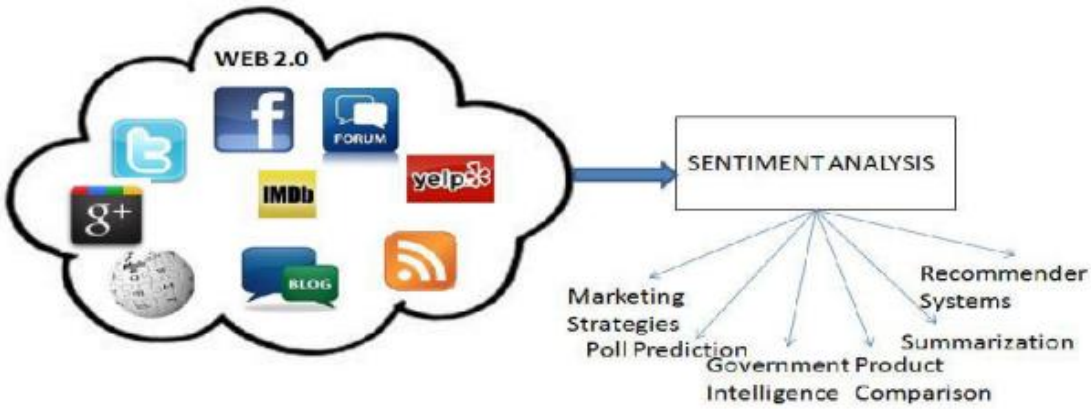


Fig 1.1 Classification of Sentiment Analysis

5.1. Social Media Monitoring

Sentiment analysis is used in social media monitoring, allowing businesses to gain insights into how customers feel about certain topics, and detect urgent issues in real-time before they spiral out of control.[29]

Example: social media monitoring: Facebook,twitter.
Twitter can evaluated if the tweets are negative or positive, as it is show on **Fig1.2.**

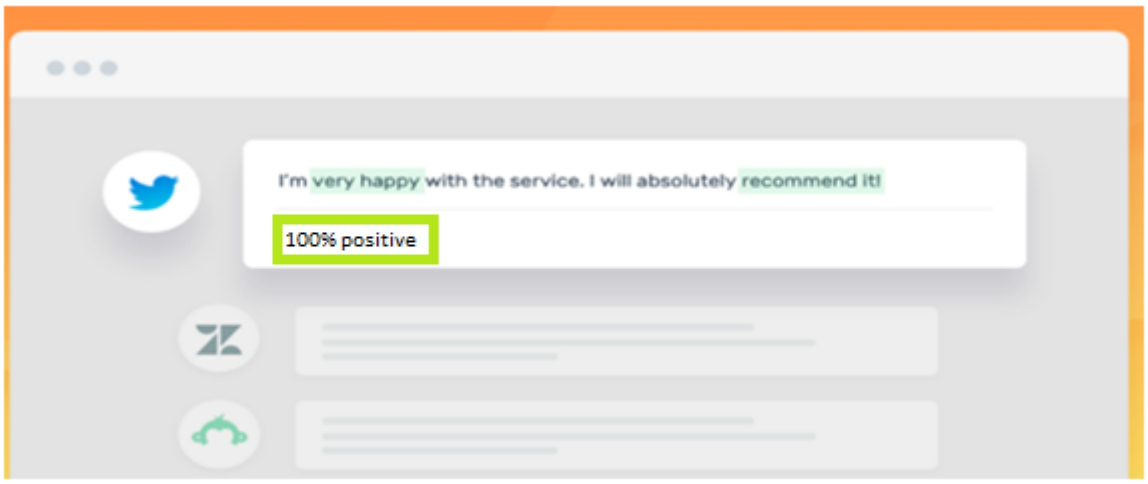


Fig 1.2 Twitter evaluat the tweets [30]

5.2. Brand Monitoring

Sentiment analysis allows you to automatically monitor all public comments regarding your brand and detect and solve potentially explosive circumstances while there is still time to defuse them.

Exemple: Hotel Sentiment

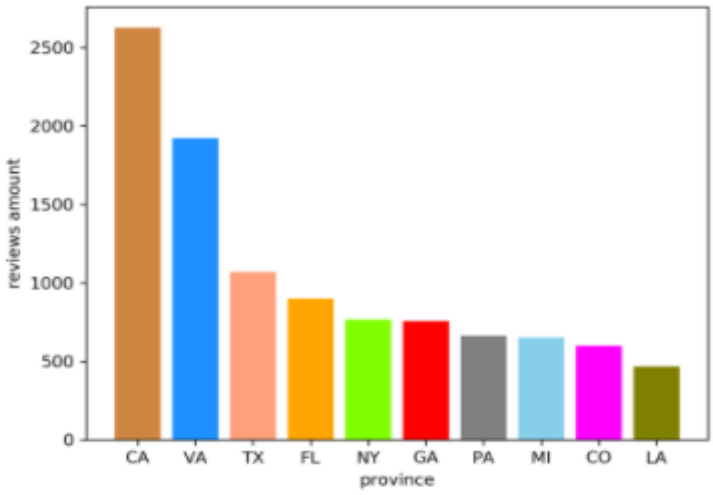


Fig1.3. The top ten provincese with most reviews [31]

5.3. Voice of the customer (VoC)

Net Promoter Score (NPS) surveys are one of the most popular ways for businesses to gain feedback with the simple question: Would you recommend this company, product, and/or service to a friend or family member? These result in a single score on a number scale.

Businesses use these scores to identify customers as promoters, passives, or detractors. The goal is to identify the overall customer experience and find ways to elevate all customers to the "promoter" level, where they, theoretically, will buy more, stay longer, and refer other customers.

5.4. Customer Service

We all know the drill: stellar customer experiences mean a higher rate of returning customers. Leading companies know that how they deliver is just as, if not more, important as what they deliver. Customers expect their experience with companies to be immediate, intuitive, personal, and hassle-free. If not, they'll leave and do business elsewhere. Did you know that one in three customers will leave a brand after just one bad experience?

You can use sentiment analysis and text classification to automatically organize incoming support queries by topic and urgency to route them to the correct department and make sure the most urgent are handled right away.

5.5. Market Research

Sentiment analysis empowers all kinds of market research and competitive analysis. Whether you're exploring a new market, anticipating future trends, or seeking an edge on the competition, sentiment analysis can make all the difference.

Example: Product Sentiment Analysis

Fig1.4. Sentimented analysis for banking-aspects[32]

6. Sentiment Analysis with movies

Big challenges can be faced in movies reviews sentiment analysis:

a) Neutral movies reviews are way more common than positive and negative ones, This is different from other sentiment analysis domains (e.g. product reviews), which tend to be predominantly positive or negative.

b) There are linguistic representational challenges, like those that arise from feature engineering issues.

c) Movies reviews are very short and often show limited sentiment cue

7. Types of Sentiment Analysis

7.1. Fine-grained Sentiment Analysis fine-grained analysis helps you determine the overall polarity of your customer reviews

7.2. Emotion detection emotion detection helps you detect emotions. This can include anger, sadness, happiness, … etc. Emotion detection systems typically use lexicons – a collection of words that convey certain emotions. Some advanced classifiers also utilize robust machine learning (ML) algorithms.

7.3. Aspect-based Sentiment Analysis aspect-based analysis go deeper. It helps you determine the particular aspects people are talking about.

7.4. Intent analysis the intent analysis helps you identify the intent of the consumer – whether the customer intends to purchase or is just browsing around.

8. Sentiment analysis algorithms

8.1. Rule-based: These systems are using a set of manually developed rules to perform sentiment analysis automatically..

8.2. Automatic: systems rely on machine learning techniques to learn from data.

8.3. Hybrid systems combine both rule-based and automatic approaches.

9. Techniques for Sentiment Analysis

Sentiment analysis relies on two types of techniques:

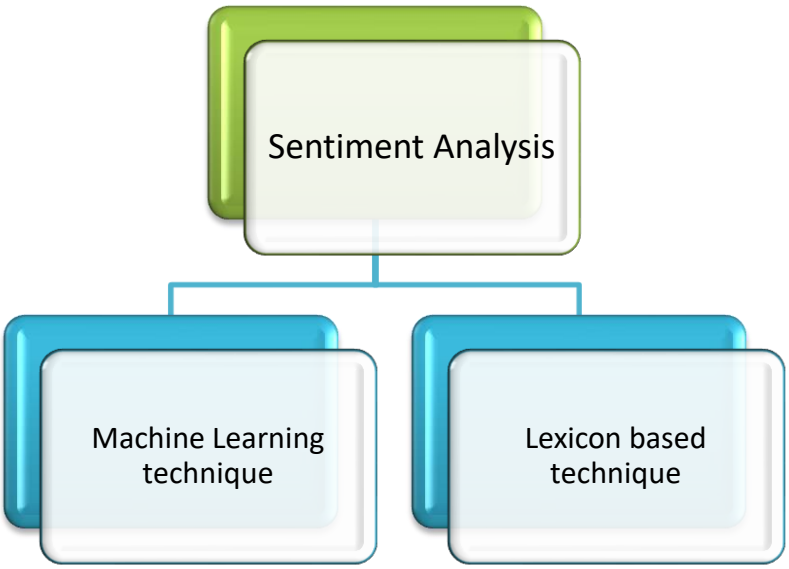


Fig1.5 Sentiment classification techniques.

a) Machine learning based technique: This type of techniques are implemented by extracting the sentences and aspect levels. The features consist of Parts of Speech (POS) tags, n-grams, bi-grams, uni-grams and bag-of-words. Machine learning contains three flavors at sentence and aspect, i.e., Naive Bayes, Support Vector Machine (SVM) and Maximum Entropy.

b) Lexicon based or corpus based technique: These techniques are based on decision trees , k-Nearest Neighbors (k-NN), Conditional Random Field (CRF), Hidden Markov Model (HMM), Single Dimensional Classification (SDC) and

Sequential Minimal Optimization (SMO), related to methodologies of sentiment classification.

Application of a lexicon is one of the two main approaches in sentiment analysis and it involves calculating the sentiment from the semantic orientation of word or phrases that occur in a text [33]. With this approach a dictionary of positive and negative words is required, with a positive or negative sentiment value assigned to each of the words. Different approaches to create dictionaries have been proposed, including manual and automatic approaches[34][35].

10. Semantic Extraction Models

- **Keyword extraction:** finding relevant words and expressions in a text. This technique is used alone or alongside one of the above methods to gain more granular insights. For instance, you could analyze the keywords in a bunch of tweets that have been categorized as "negative" and detect which words or topics are mentioned most often.[36]
- **Entity extraction:** identifying named entities in text, like names of people, companies, places, etc. A customer service team might find this useful to automatically extract names of products, shipping numbers, emails, and any other relevant data from customer support tickets.

Automatically classifying tickets using semantic analysis tools relieves agents from repetitive tasks and allows them to focus on tasks that provide more value while improving the whole customer experience.

Tickets can be instantly routed to the right hands, and urgent issues can be easily prioritized, shortening response times, and keeping satisfaction levels high.

11. Elements of Semantic Analysis

The bellow are some important elements of semantic analysis :

11.1. Hyponymy

It may be defined as the relationship between a generic term and instances of that generic term. Here the generic term is called hypernym and its instances are called hyponyms. For example, the word color is hypernym, and the color blue, yellow, etc. are hyponyms.

11.2. Homonymy

It may be defined as the words having the same spelling or same form but having different and unrelated meanings. For example, the word “Bat” is a homonymy word because a bat can be an implement to hit a ball or a bat is a nocturnal flying mammal also.

11.3. Polysemy

Polysemy is a Greek word, which means “many signs”. It is a word or phrase with a different but related sense. In other words, we can say that polysemy has the same spelling but different and related meanings[37]. For example, the word “bank” is a polysemy word having the following meanings :

- A financial institution.
- The building in which such an institution is located.
- A synonym for “to rely on”.

11.4. Difference between Polysemy and Homonymy

Both polysemy and homonymy words have the same syntax or spelling.[26] The main difference between them is that in polysemy, the meanings of the words are related but in homonymy, the meanings of the words are not related. For example, if we talk about the same word “Bank”, we can write the meaning ‘a financial institution’ or ‘a river bank’. In that case, it would be an example of a homonym because the meanings are unrelated to each other.

11.5. Synonymy

It is the relation between two lexical items having different forms but expressing the same or a close meaning. Examples are ‘author/writer’, ‘fate/destiny.

11.6. Antonymy

It is the relation between two lexical items having symmetry between their semantic components relative to an axis. The scope of antonymy is as follows :

- Application of property or not − Example is ‘life/death’, ‘certitude/incertitude’
- Application of scalable property − Example is ‘rich/poor’, ‘hot/cold’
- Application of a usage − Example is ‘father/son’, ‘moon/sun

12. Building Blocks of Semantic System

In word representation or representation of the meaning of the words, the following building blocks play an important role

- **Entities** : It represents the individual such as a particular person, location, etc. For example, Haryana. India, Ram all are entities.
- **Concepts :** It represents the general category of the individuals such as a person, city, etc.
- **Relations :** It represents the relationship between entities and concepts. For example, Ram is a person.
- **Predicates :** It represents the verb structures. For example, semantic roles and case grammar are examples of predicates.[38]

13. Conclusion

In this chapter , we learned about the importance of sentiment analysis, and we talked about of all principale points , as well as analysis sentiment with movies and its related works .

.

CHAPTER 2

MACHINE LEARNING AND DEEP LEARNING

1. Introduction:

Artificial intelligence (AI) has received a lot of media attention in recent years. Innumerable articles mention machine learning, deep learning, and AI.

The essential ideas and concepts of machine learning and deep learning, as well as the differences between them, are introduced in this second chapter.

Figure2.1. depicts Artificial Intelligence paradigms.

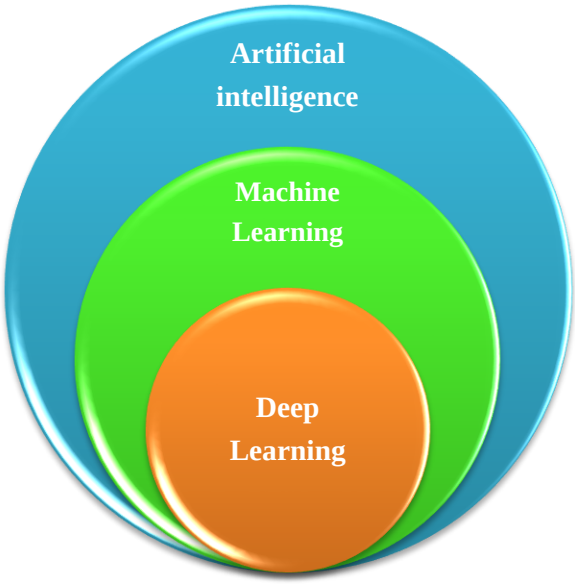


Fig 2.1: Artificial Intelligence paradigms

2. Machine Learning:

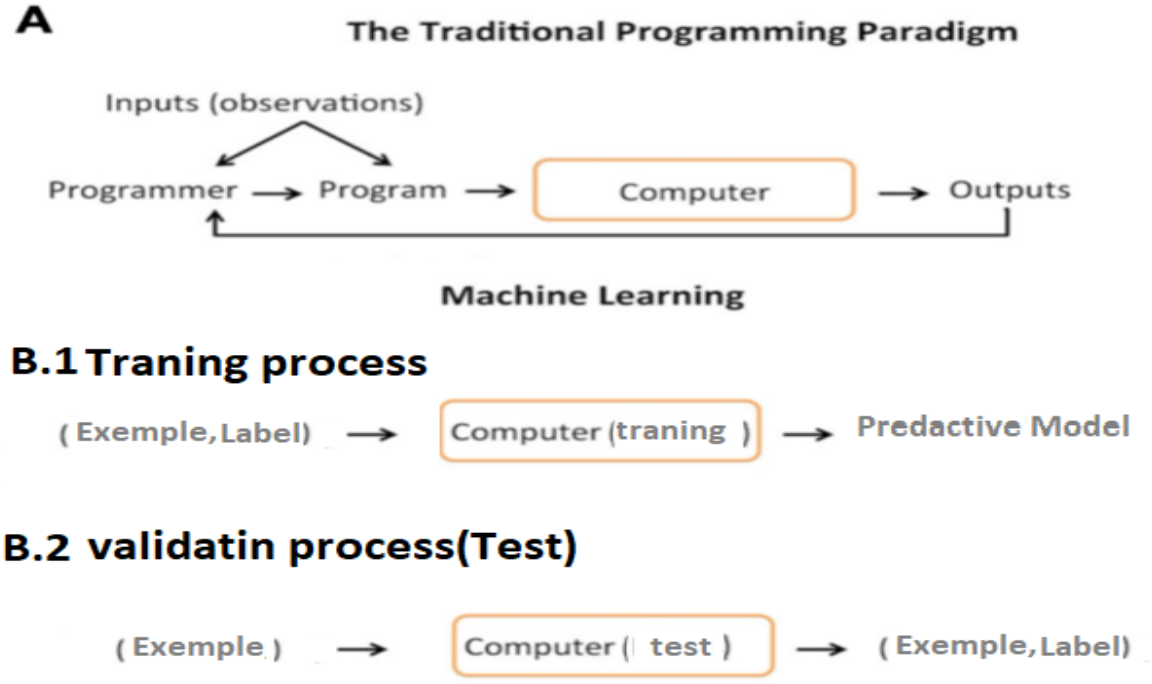


Fig 2.2. Comparison between traditional programming (A) and machine learning (B)

2.1. Definition:

Machine learning (ML) is a type of a AI that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. [39]

2.2. Machine Learning types: There are four basic approaches:

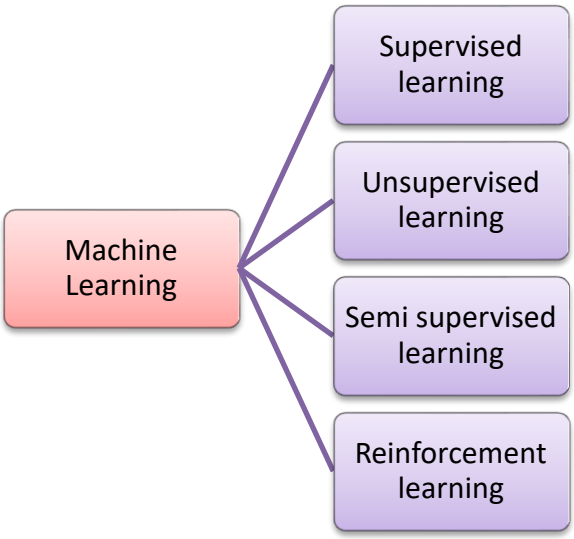


Fig 2.3 Machine Learning techniques.

2.2.1. Supervised learning

Supervised learning is concerned with predicting a target value given input observations. In machine learning, we call the model inputs "features". The target values that supervised models are trained to predict are also often called labels.[40][41]

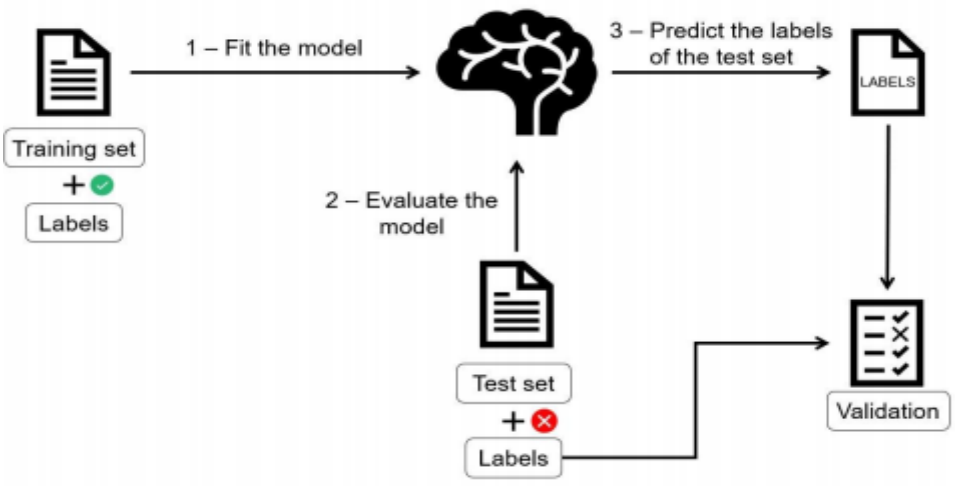


Fig.2.4 Supervised learning [42]

2.2.2. Supervised Learning Algorithms

- **Decision Trees (DTs)**: are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.[43]

A tree can be seen as a piecewise constant approximation.

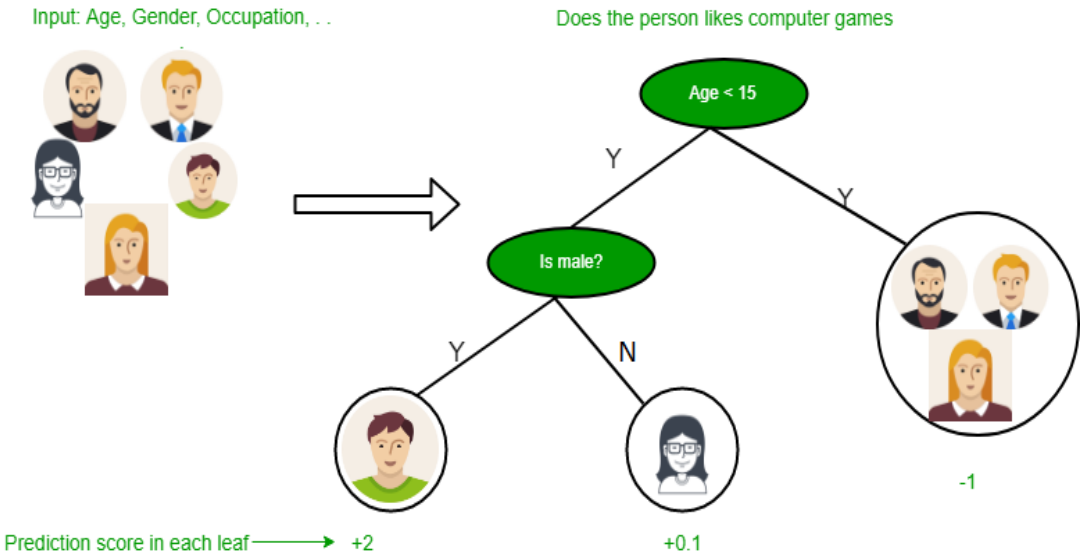


Fig 2.5 Decision tree example.[44]

- **Logistic regression:** Machine learning has also adopted logistic regression as a statistical technique. It's the approach of choice for dealing with binary classification difficulties (problems with two class values 0 and 1).

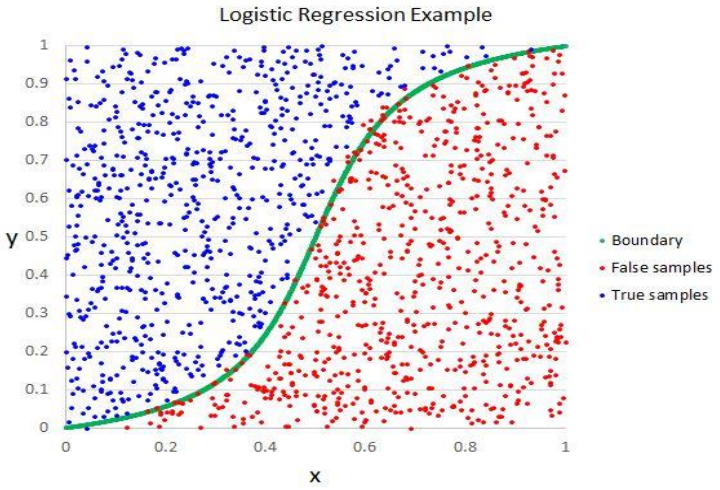


Fig 2.6 Logistic Regression example.[45]

- **Support Vector Machines (SVM)** in machine learning is a supervised learning model with the related learning algorithm, which examines data and identifies patterns, which is used for regression and classification analysis Cortes & Vapnik (1995). Recently, many classification algorithms have been proposed, but SVM is still one of the most widely and most popular used classifiers.[46] [47]

Mathematically, a separating hyperplane can be written as:

W·X+b=0, where W is a weight vector and W=w 1,w2,...,wn. X is a training tuple. b is a scalar.If the data is linearly inseparable, the SVM uses nonlinear mapping to transform the data into a higher dimension. It then solve the problem by finding a linear hyperplane. Functions to perform such transformations are called kernel functions. [48]

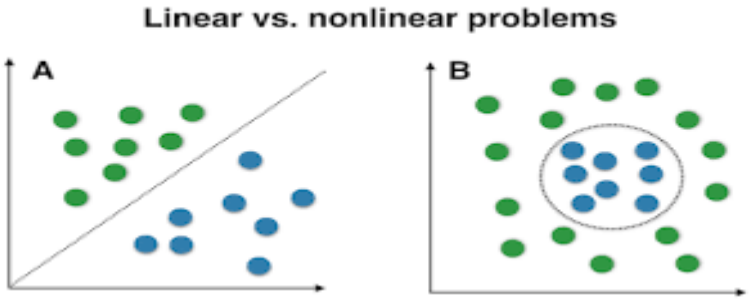


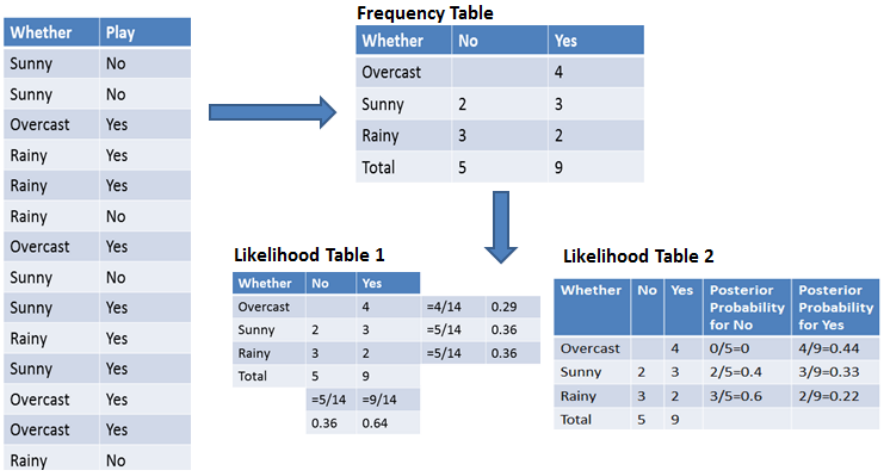
Fig 2.7 Nonlinear and linear SVM classification example.[49]

- **Naïve Bayes:** Naive Bayes is a simple supervised machine learning algorithm that uses the Bayes' theorem with strong independence assumptions between the features to procure results. That means that the algorithm just assumes that each input variable is independent. It really is a naive assumption to make about real-world data[50]. It is based on the application of the Baye's rule given by the following formula:

$$P(C = c/D = d = \frac{P(D = d/C = c)P(C = c)}{P(D = d)} \qquad (1)$$

where D denotes the document and C the category (label), d and c are instances of D and C and P(D = d) = ∑ (D |C)P(C). We can simplify this expression by,

$$P(c/d) = \frac{P(d/c)P(c)}{P(d)} \quad (2)$$



Whether	Play
Sunny	No
Sunny	No
Overcast	Yes
Rainy	Yes
Rainy	Yes
Rainy	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table

Whether	No	Yes
Overcast		4
Sunny	2	3
Rainy	3	2
Total	5	9

Likelihood Table 1

Whether	No	Yes		
Overcast		4	=4/14	0.29
Sunny	2	3	=5/14	0.36
Rainy	3	2	=5/14	0.36
Total	5	9		
	=5/14	=9/14		
	0.36	0.64		

Likelihood Table 2

Whether	No	Yes	Posterior Probability for No	Posterior Probability for Yes
Overcast		4	0/5=0	4/9=0.44
Sunny	2	3	2/5=0.4	3/9=0.33
Rainy	3	2	3/5=0.6	2/9=0.22
Total	5	9		

Fig 2.8 Naïve Bayes classification example.[51]

- **K-nearest neighbor algorithm:**

  A k-nearest-neighbor algorithm, often abbreviated k-nn, is an approach to data classification that estimates how likely a data point is to be a member of one group or the other depending on what group the data points nearest to it are in [52].

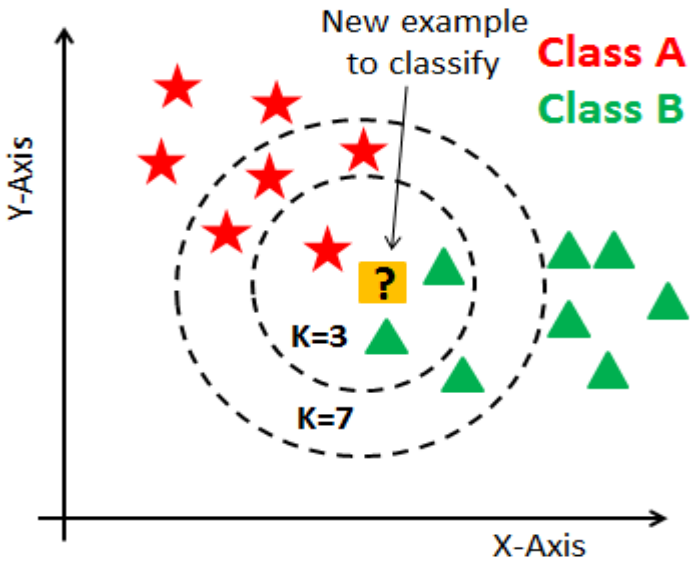


Fig 2.9 KNN classification example.[53]

2.2.2.1. Application of Supervised Machine Learning

1- Bioinformatics
2- Quantitative structure
3- Database marketing
4- Handwriting recognition
5- Information extraction
6- Object recognition in computer vision
7- Optical character recognition
8- Spam detection
9- Pattern recognition

2.2.3. Unsupervised learning

The training set is left unlabeled in unsupervised learning. As a result, this strategy is more suitable for dealing with clustering problems.[26] This sort of machine learning is extremely useful when the goal is to obtain a better grasp of how data is organized in order to uncover possible patterns. Algorithms with no supervision

- It algorithms is :
  - K-Means Clustering
  - Gaussian Mixture Model
  - Hidden Markov Model
  - Principal Component Analysis (PCA)
- Unsupervised Machine Learning Application:
  - Human Behavior Analysis
  - Social Network Analysis to define groups of friends.
  - Market Segmentation of companies by location, industry, vertical.
  - Organizing computing clusters based on similar event patterns and processes.

2.2.4. Semi-supervised learning

Semi-supervised learning is supervised learning where the training data contains very few labeled examples and a large number of unlabeled examples.The goal of a semi-

supervised learning model is to make effective use of all of the available data, not just the labelled data like in supervised learning.[54]

In semi-supervised learning we are given a few labeled examples and must make what we can of a large collection of unlabeled examples. Even the labels themselves may not be the oracular truths that we hope for.[55]

2.2.5. Reinforcement learning

Reinforcement Learning is a type of Machine Learning which allows machines to automatically determine the ideal behavior within a specific context, in order to maximize its performance. The reinforcement learning method aims at using observations gathered from the interaction with the environment to take actions that would maximize the reward or minimize the risk. To produce intelligent programs (also called agents).[56]

3. Deep Learning

3.1. Definition

Deep learning is a process that allows computers to learn to execute activities that are inherent in the brain, such as picture identification, as a new branch of machine learning research.

Today, the deep learning approach DL is the new trend in machine learning, as it provides far more advanced pattern recognition and picture categorization than the old machine learning approach ML.[57]

3.2. Deep Neural Network

An artificial neural network (ANN) is a form of artificial intelligence that aims to replicate the learning process that people utilize to acquire specific types of knowledge[56]. Artificial neurons, like biological neurons in the brain, are present in ANN, and they are used to recognize and retain information. An ANN is made up of input and output layers, as well as one or more hidden layers (in most situations).[58][59]

3.2.1. Activation function

It’s just a thing function that you use to get the output of the node. It is also known as Transfer Function.

3.2.2. Activation function types

- ***Sigmoid:*** is an adjective that designates something in the shape of an "S".

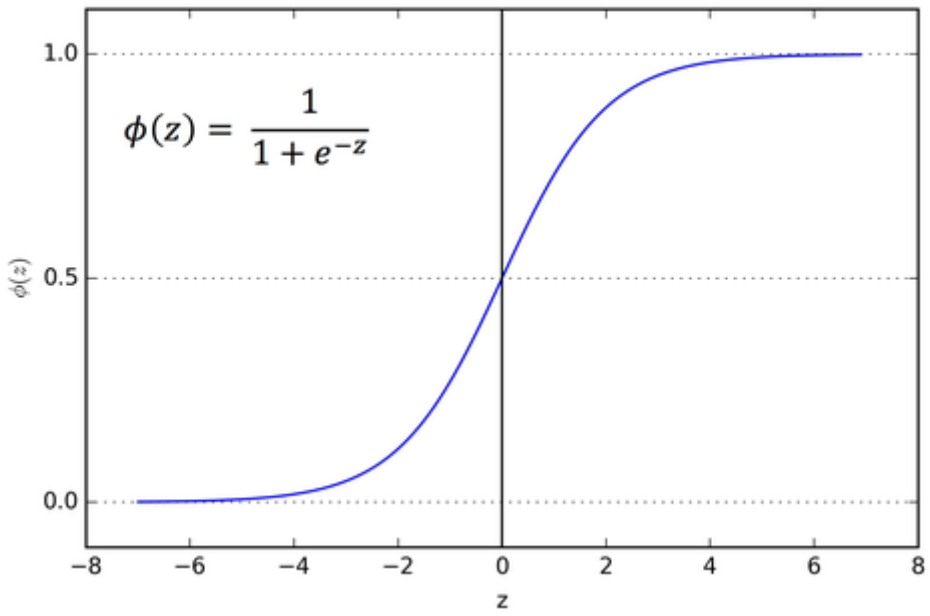


Fig 2.10 Sigmoid function curve[60]

- ***ReLU:*** The Rectified Linear Unit is the most commonly used activation function in deep learning models. The function returns 0 if it receives any negative input, but for any positive value xx it returns that value back. So it can be written as f(x)=max(0,x)f(x)=max(0,x).

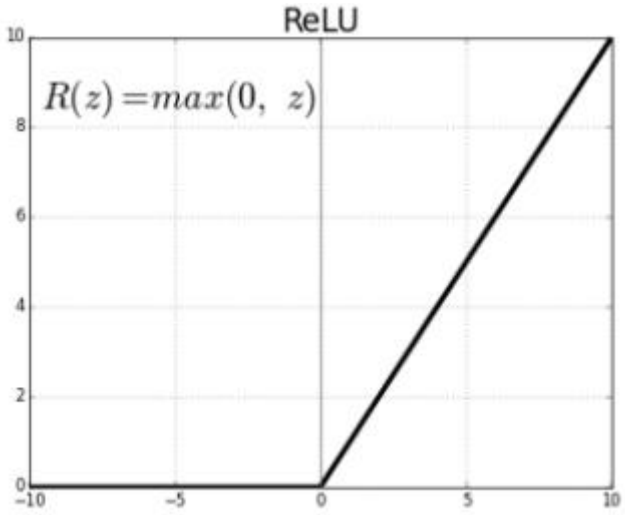


Fig 2.11 ReLU function [60]

- ***SoftMax:*** The softmax function is a function that turns a vector of K real values into a vector of K real values that sum to 1. The input values can be positive, negative, zero, or greater than one, but the softmax transforms them into values between 0 and 1 so that they can be interpreted as probabilities. If one of the inputs is small or negative, the softmax turns it into a small probability, and if the input is large, then it turns it into a large probability, but it will always remain between 0 and 1.
- **TanH**: tanh is also like logistic sigmoid but better. The range of the tanh function is from (-1 to 1). tanh is also sigmoidal (s-shaped).

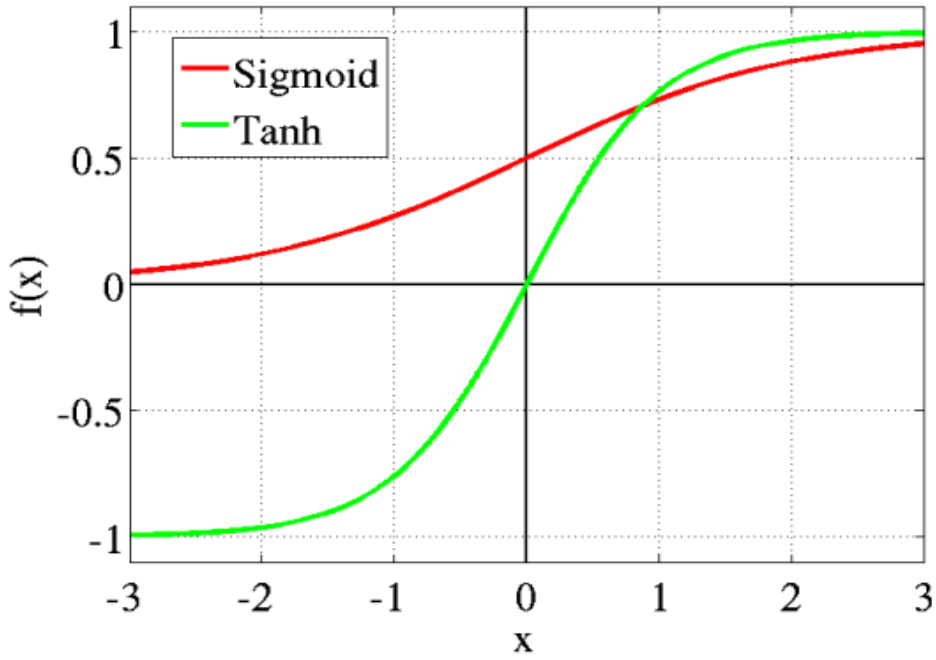


Fig 2.12 Tanh v/s Logistic Sigmoid [60]

3.2.3. Weights and Bias

Weight is a number that represents the strength of a connection between two nodes; weights are what connect nodes between layers; weights are randomly initialized (numbers) with a mean of 0 and a standard deviation of 1, and each neuron has its own bias. These biases are learnable, it is updates the weights, it also updates the bias. The bias determines whether or not the neuron will fire through the network (forward pass), and the bias is passed to the activation function with the SUM of the weights.

4. Deep Neural Network types

4.1. Deep Feed-Forward Neural Network

Multi-layered Network of neurons is composed of many sigmoid neurons. MLNs are capable of handling the non-linearly separable data. The layers present between the input and output layers are called hidden layers. The hidden layers are used to handle the complex non-linearly separable relations between input and t output.

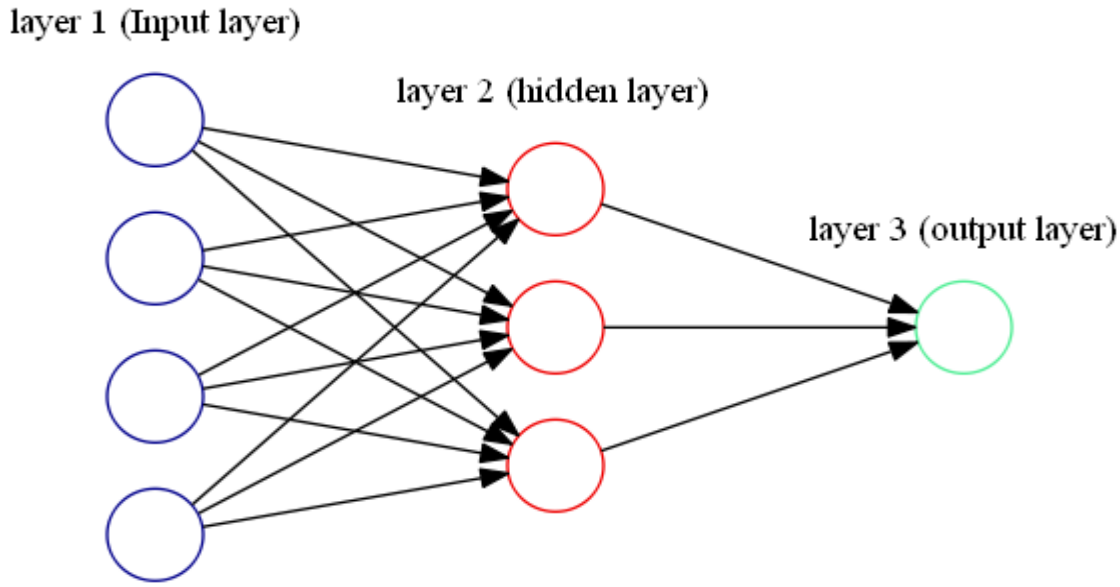


Fig 2.13 Representation of a simple feedforward neural network.[61]

4.2. Deep Feed-Forward Neural Network layers

- *Input layer:* The input layer is the first layer in the neural network, composed of input neurons, and brings initial data to the hidden layers for further processing.[62]

- *Hidden layer:* In neural networks, a hidden layer is located between the input and output of the algorithm.In which the function applies weights to the inputs and directs them through an activation function as the output. In short, the hidden layers perform nonlinear transformations of the inputs entered into the network. Hidden layers vary depending on the function of the neural network, and similarly, the layers may vary depending on their associated weights.[63]

- ***Output layer:*** The output layer is the last in a neural network that produces the outputs of the program, in classification tasks, the size of the output layer is equal to many classes.

4.3. Deep Convolutional Neural Network

A convolutional neural network (or CNN) is a special type of multilayer neural network or deep learning architecture inspired by the visual system of living beings. The CNN is very much suitable for different fields of computer vision and natural language processing.[64]

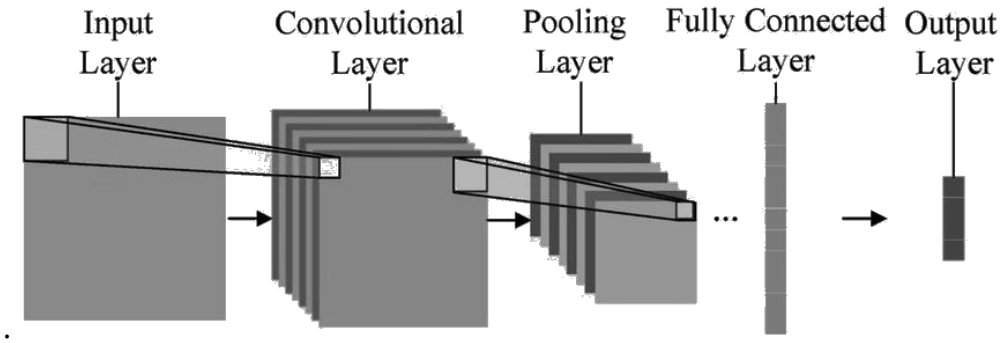


Fig 2.14 Convolutional Neural Network architecture.[65]

4.3.1. Deep convolutional Neural Network layers

There are three main types of layers to build Convolutional Neural Network architectures: Convolutional Layer, Pooling Layer, and Fully-Connected Layer.

- ***input***

  Images are sent into CNNs. Images are represented as a three-dimensional array of pixel values (height, width, and RGB values (color channels)). An image with height = 480 and width = 480, for example, will be shown as a $480 \times 480 \times 3$ tensor..

- **Convolutional Layer (CONV layers)**

  In a CNN, this sort of layer is always the initial layer. Filters (kernels) that span all regions of the input image are applied to CONV layers. The receptive field is the region that the filter covers. The filter's dimensions are height, breadth, and depth. Where the layer's input is the same as the depth.

  The values in the filter are multiplied (element-wise multiplication) with the original pixel values of the picture as the filter strides (convolves) through the input image. For each place in the input volume, the operation is repeated. The stride option specifies the number of units by which the filter is shifted. A

feature or activation map is the end product. The deeper the activation map and the more information about the input image, the more filters there are.[66]

- ***Pooling Layer:***

  In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon the method used, there are several types of Pooling operations.

  In Max Pooling, the largest element is taken from the feature map. Average Pooling calculates the average of the elements in a predefined sized Image section. The total sum of the elements in the predefined section is computed in Sum Pooling. The Pooling Layer usually serves as a bridge between the Convolutional Layer and the FC Layer [67].

- ***Fully-Connected Layer:***

  Different layers in a neural network are connected. In some cases, all neurons in layer X are connected to neurons in the next hidden layer X+1. This is called a fully connected layer, usually followed by a nonlinear activation function.[62]

4.4. Deep Recurrent Neural Network:

A recurrent Neural Network is a generalization of a feedforward neural network that has internal memory. RNN is recurrent as it performs the same function for every input of data while the output of the current input depends on the past computation. After producing the output, it is copied and sent back into the recurrent network. For making a decision, it considers the current input and the output that it has learned from the previous input.

Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. In other neural networks, all the inputs are independent of each other. But in RNN, all the inputs are related to each other.

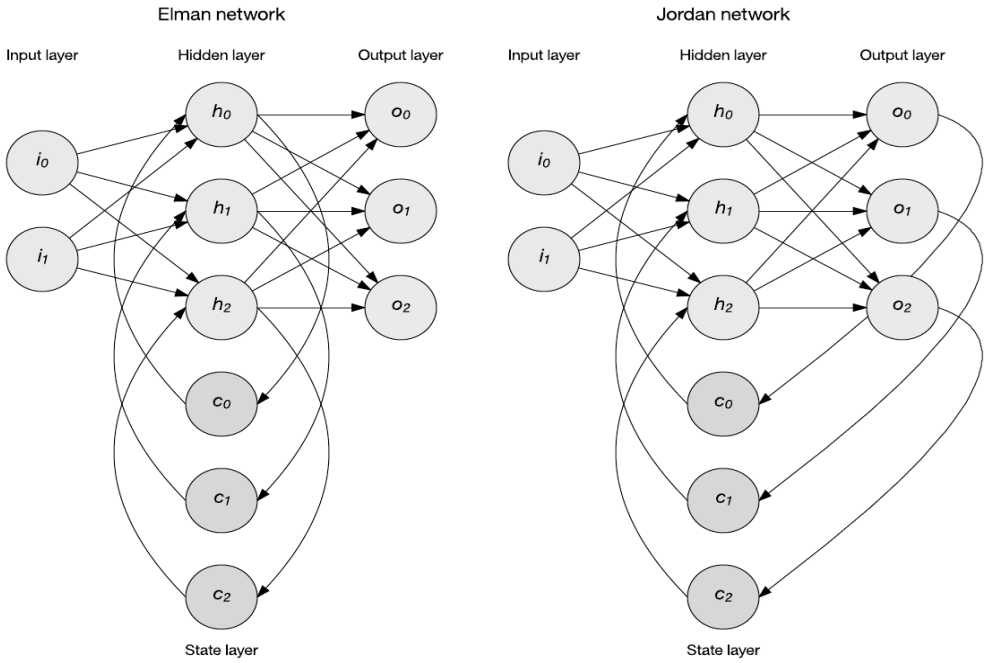


Fig 2.15 Recurrent Neural Network architecture.[68]

5. ML vs DL:

	Machine Learning	**Deep Learning**
Approach	Machine learning uses algorithms to parse data, learn from that data, and make informed decisions based on what it has learned	Deep learning structures algorithms in layers to create an "artificial neural network" that can learn and make intelligent decisions on its own
Time	Takes less time to train	Takes a long time to train
Hardware	Trains on CPU	Trains on GPU for proper training
output	The output is in numerical form for classification and scoring applications	The output can be in any form including free form elements such as free text and sound
Capability of training	Can train on lesser training data	Requires large data sets for training
Human Intervention	a human needs to identify and hand-code the applied features based on the data type	tries to learn those features without additional human intervention.

Table 2.1 The differences between ML and DL

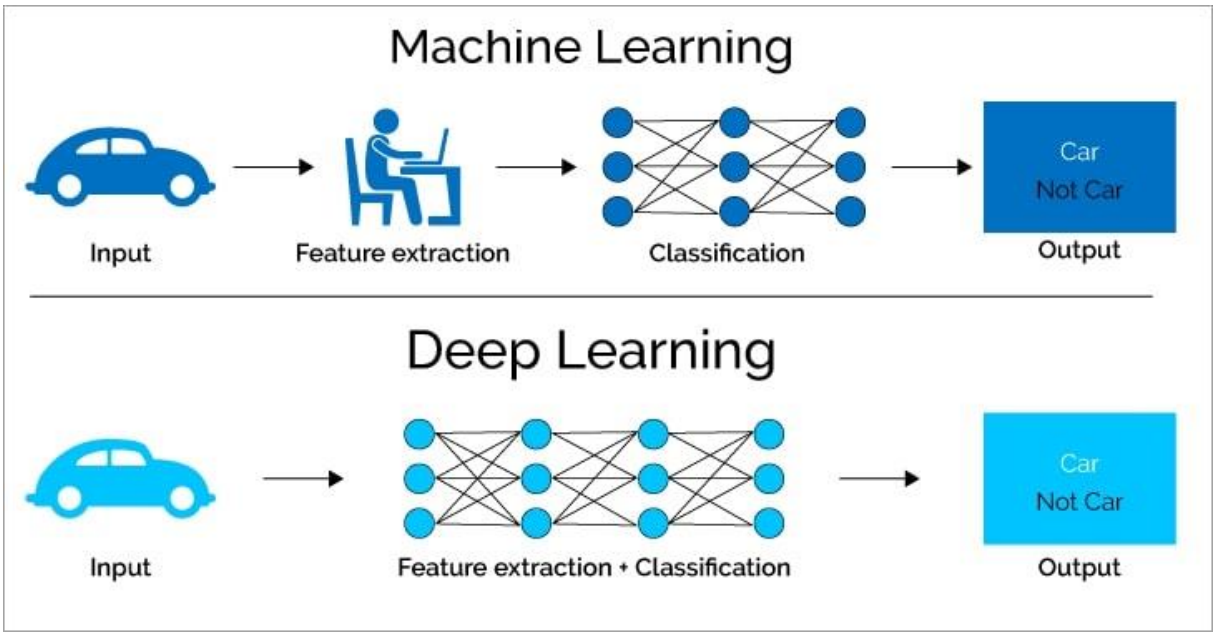


Fig 2.16 Difference between ML and DL[69]

6. Conclusion

This chapter introduced the main ideas (or desires) of machine learning and deep learning and explained the relationship between them. In sum, the takeaway is that deep learning is a subfield of machine learning, and artificial intelligence and machine learning are not synonyms. Moreover, this chapter introduced the three broad categories of machine learning, that is, working with labeled data (supervised learning) and unlabeled data (unsupervised learning), and learning complex processes (reinforcement learning). Ultimately, there is a lot of domain-specific jargon surrounding deep learning. A lot of technical terms are from other fields, but be aware that familiar terms may refer to something different in the context of deep learning.

CHAPTER 3

ENVIRONMENT TOOLS

1. Introduction

After having studied in the second part AI, ML and DL.We will present in this part an in-depth study on the programming environment and the tools used.

2. Tools presentation

2.1. Hardware

Deep learning is an area with high computational needs and the availability of resources dedicated to this task will fundamentally influence the user experience because without these resources.The experiments have all been carried out on a machine that offers acceptable performance whose characteristics are:

- A laptop PC HP i5-6200U CPU @ 2.30GHz 2.40 GHz
- RAM size 8GO
- Hard drive size 500 GB
- Intel(R) HD Graphics 520 Card
- 64-bit Windows 10 operating system

2.2. The used tools and libraries

2.2.1. Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components. Python's simple, easy-to-learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed. [70]

2.2.2. ANACONDA

Anaconda is a packaged compilation of Python and a set of packages containing various libraries, including core libraries that are widely used in data science. The main advantage of this release is that we don't need complicated settings, and can run on all types of operating systems and platforms (especially Windows), which usually causes problems when installing specific Python packages. Therefore, we only need to download and install it once to start our data science journey. Anaconda distributions are widely used in data science environments throughout the industry. [71]

Anaconda comes with a wonderful IDE, Spyder (Scientific Python Development Environment), besides other useful utilities like Jupyter notebooks, the IPython console, and the excellent package management tool, conda that let us install, remove, or upgrade any Anaconda package with a single command in Anaconda Prompt. In our experimental work, we use Anaconda 5.0.1 which is compatible with python 3.6. [72]

Fig 3.1. Logo of Anaconda

2.2.2.1. Anaconda installation

A. Download

To download Anaconda 5.0.1 for windows10 / 64x:

- Click on this link https://repo.anaconda.com/archive/.
- Choose the anaconda3-5.0.1 from the list as indicated in Fig 3.2.

Fig 3.2 Anaconda 5.0.1 downloading. [73]

B. Installation

- Search in Google " Anaconda Python "➔ Python V 3.6 for Windows 64 bits
- Install the Anaconda 64-bit installation file for Windows
  Install for ➔ option: Just for me
- Add the installation path to the variables environment (Heading:
  PATH) to facilitate access
- After the installation is finished, open Anaconda Prompt (C:\Users\....\Anaconda3)
- Open Anaconda Prompt in ADMINISTRATOR mode. This will avoid errors in writing rights later when installing libraries/packages. (anaconda prompt ➔ right button ➔ menu ➔ administrator mode ➔ C:\WINDOWS\system32>
- Type the command for install libraries

C. Python libraries

- **sklearn**

Formerly sci-kit-learn and also known as sklearn; Is a free software machine learning library, and it's one of the most important and indispensable Python frameworks for Data Science and Machine Learning in Python. It applies a wide range of machine learning algorithms that cover the main areas of machine learning like classification, clustering, regression, and so on. All the mainstream Machine Learning algorithms like support vector machines, logistic regression, random forests, K-means clustering, hierarchical clustering, and many more. In this work, we have installed Sklearn 0.19.1.[74]

To install sklearn

1- Anaconda prompt.

2- Type the command : python -m pip install --upgrade sklearn== 0.19.1

- **SciPy**

SciPy (pronounced “Sigh Pie”) is a Python-based ecosystem of open-source software for math ematics, science, and engineering. In particular, these are some of the core packages:. In this work, we have installed SciPy 1.2.1.[75]

To install SciPy

1- Anaconda prompt.

2- Type the command: python -m pip install --upgrade scipy == 1.2.1

- **Pandas**

Pandas is a popular Python-based data analysis library, it presents a diverse range of utilities, ranging from parsing multiple file-formats to converting an entire data table into a NumPy ma trix array. This makes pandas a trusted ally in data science and machine learning. In this work, we have installed pandas 0.24.1. [76]

To install pandas:

1- Anaconda prompt.

Type the command: python -m pip install --upgrade pandas ==24.1.

- **NumPy**

NumPy stands for "Numerical Python" is an open-source package for the Python programmin g language. It is a library consisting of multidimensional array objects, along with a large coll ection of high-level mathematical functions to operate on these arrays. In our work, we have i nstalled NumPy 1.16.1. [77]

To install NumPy:

1- Anaconda prompt.

2- Type the command: python -m pip install --upgrade numpy==1.16.1.

- **Matplotlib**

Matplotlib is a Python package, it supports interactive and non-interactive 2D plotting that generates production-quality graphs, and can save images in several output formats. It provides various types of plots (lines, bars, pie charts, histograms...etc.). In addition to this, it is flexible and easy to use. In this work, we have installed Matplotlib 2.2.2. [78]

To install Matplotlib:

1- Anaconda prompt.

2- Type the command pip install matplotlib.

- **TensorFlow**

TensorFlow is an open-source library for fast numerical computing which implements machine learning methods based on deep neural networks (deep learning). At a high level, TensorFlow is a Python library that allows users to express arbitrary computation as a graph of data flows. We have installed TensorFlow 1.2.1. [79]

To install TensorFlow:

1- Anaconda prompt.

2- Type the command: python -m pip install --upgrade tensorflow==1.2.1

- **Keras**

Keras is a deep learning API written in Python, and its development purpose is to make developing DL models as fast and easy as possible. Keras is used as the backend TensorFlow and Theano. In the implementation part of this work, we have used Keras 2.2.4 with the TensorFlow backend. [80]

To install Keras:

1- Anaconda prompt.

2- Type the command: python -m pip install --upgrade keras == 2.0.6

- **JSON**

JSON (JavaScript Object Notation), is a lightweight data-interchange format inspired by JavaScript object literal syntax (although it is not a strict subset of JavaScript).

JSON exposes an API familiar to users of the standard library marshal and pickle modules. In this work ,we have installed JSON 2.0.9 [81]

To install JSON:

1- Anaconda prompt.

2- Type the command: python -m pip install --upgrade json== 2.0.9

- **CSV**

A Comma Separated Values (CSV) file is a plain text file that contains a list of data. These files are often used for exchanging data between different applications. For example, databases and contact managers often support CSV files. [82]

- **GaussianNB**

GaussianNB classifiers, also known as Naive Bayes classifiers, are a collection of supervised learning algorithms based on Bayes' theorem, but with strong independence assumptions between the features given the class variable's value (hence naive).

- **Itertools**

Python’s itertools library is a gem , you can compose elegant solutions for a variety of problems with the functions it provides. In more-itertools we collect additional building blocks, recipes, and routines for working with Python iterables.

3. Dataset Description

We use the IMDB Dataset, which is publicly available on Kaggle website. Here is a description of the data:

We have three files : IMDB Dataset.csv to traning data, Test.csv, and Valid.csv to testing.

We used a 50000 reviews for training , while 5000 reviews for the test model.

Variable	Description
review	description of the review posted
sentiment	target variable

Table 3.1 Description of IMDB Dataset

We are interested only in 2 columns. ‘review’ contains movies reviews given by different users and ‘sentiment’ which keeps the record of ‘Positive’ or ‘Negative’. So, in essence, this is simply a 2-class sentiment analysis problem.

- we applied two kinds of features: Positive and Negative
- ‘Positive’ is a word that has a significant impact on identifying the polarity of opinion positive.
- ‘Negative’ is a word that has a significant impact on identifying the polarity of opinion negative.

4. Conclusion

In this chaptre, we have presented all the environmental tools that have been used in our work. Python is a high-level programming language and the most commonly used language in data science and deep learning, which is why we chose it with their libraries.

CHAPTER 4

THE REALIZED WORK AND OBTAINED RESULTS

1. Introduction

This chapter presents the experimental part, so it contains the following:

- The preprocessing of the used dataset
- The proposed model
- Our application interface.

2. Preprocessing of the used dataset

The framework of the proposed method as shown in Fig4.1 consists of six main phases namely; Dataset, Transformation, Pre-processing, Word Vector Representation, Classification, and Evaluation. Dataset phase aims to clarify the dataset by determining the details of its content Fig 4.2:

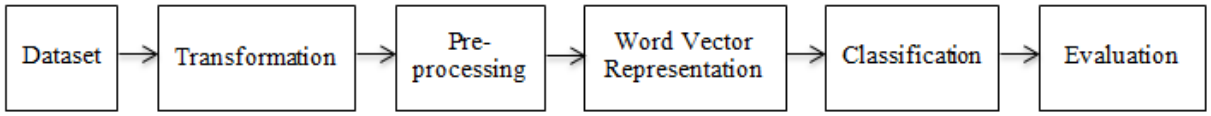


Fig 4.1. Steps for processing the reviews

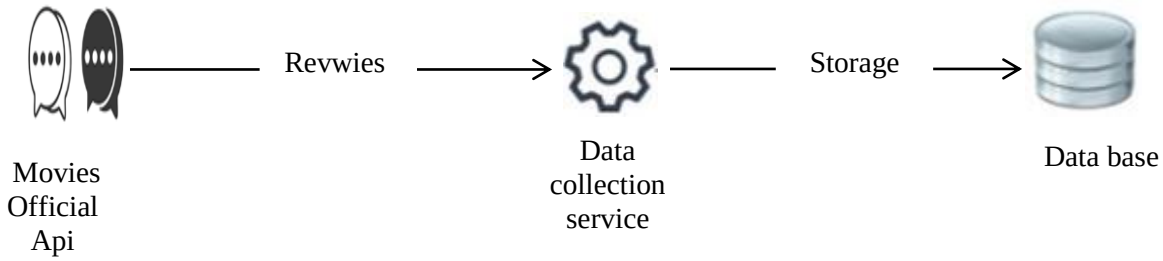


Fig 4.2. Details of the data collection flow.

```
dataset= pd.read_csv("./train.csv")
dataset1 = pd.read_csv("./test1.csv")
```

Fig 4.3 Application of load_data function source code

3. ANN model

3.1. The proposed model

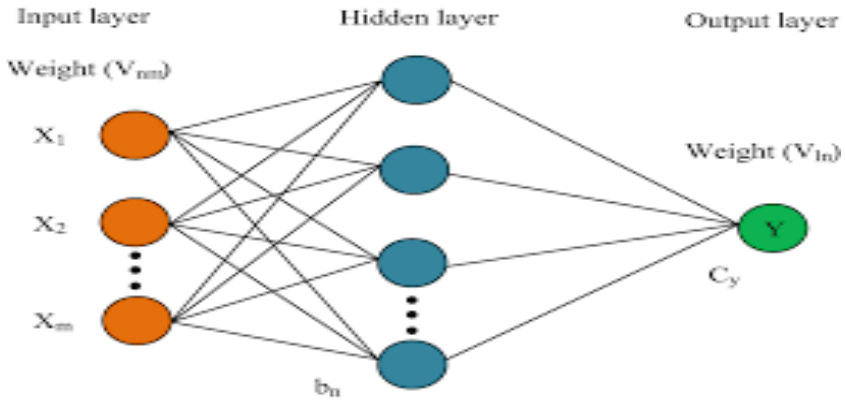


Fig 4.4 Presents a detailed architecture of the proposed ANN model.

1- Build the Model

Table 4.1: Description of the Proposed ANN Model

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 750)	7500750
batch_normalization_1 (Batch	(None, 750)	3000
dropout_1 (Dropout)	(None, 750)	0
dense_2 (Dense)	(None, 512)	384512
batch_normalization_2 (Batch	(None, 512)	2048
dropout_2 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 128)	65664
batch_normalization_3 (Batch	(None, 128)	512
dropout_3 (Dropout)	(None, 128)	0
dense_4 (Dense)	(None, 64)	8256
batch_normalization_4 (Batch	(None, 64)	256
dropout_4 (Dropout)	(None, 64)	0
dense_5 (Dense)	(None, 16)	1040

Table 4.1 summarizes the obtained results after applying the proposed ANN model

2- Train the Model

```
# Training the Model
hist=model.fit(train_X, labels,batch_size=32, epochs=15,verbose=1,validation_split=0.1,shuffle=True)
```

Fig 4.5 Represented the trained model

3.2. The obtained Results

		Loss value	**Accuracy value**
ANN model	Training set	2.58%	97.42%
	Valid set	2.58%	97.42%
Gaussian-NB	Valid set	81.86%	18.14%

Table 4.2: Loss value and accuracy value obtained.

Fig 4.6 shows the evaluation of training loss and validation loss over time in function of the number of epochs. Fig 4.7 plots the evolution of training accuracy and validation accuracy in the function of the number of epochs.

Fig 4.6 Training loss Vs Validation loss of the ANN model

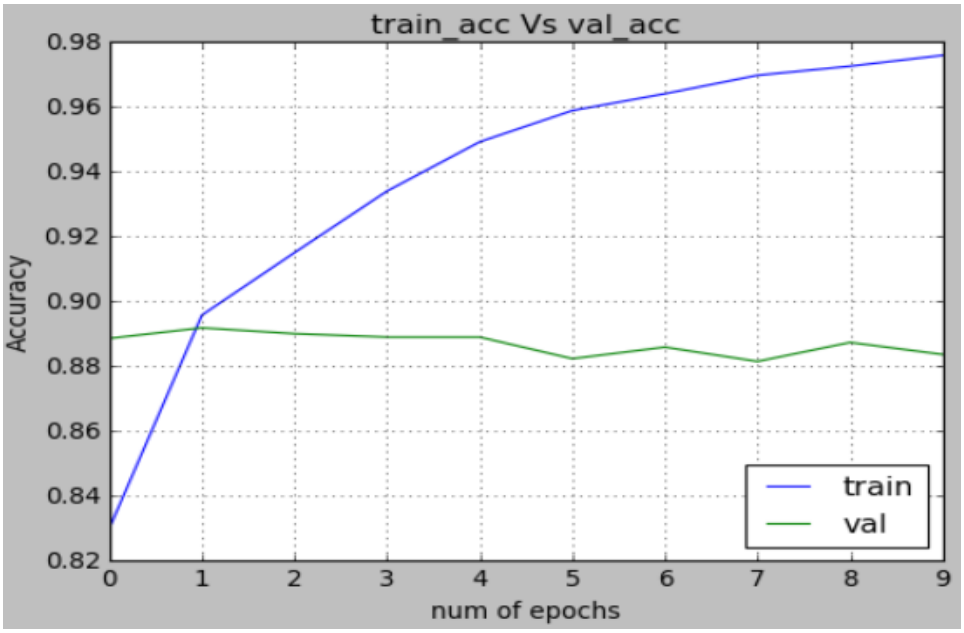


Fig 4.7 Training accuracy Vs Validation accuracy of the ANN model

3.3. The evaluation of the model :

- ***Accuracy***

Accuracy is a classification model assessment statistic that indicates how accurate the model is. Accuracy: the ratio of correctly predicted comments to the total comments. For example, the accuracy is 0.97, or 97.42% (97 correct predictions out of a total of 100 examples).[24]

- ***Loss***

The loss function is used to optimize the parameter values in the neural network model. The loss function maps a set of parameter values of the network to scalar values, which indicate the degree to which these parameters complete the tasks that the network intends to complete. It calculates the error for a single training example.[83]

- ***Optimizer***

An optimizer is an algorithm or method used to change neural network properties (such as weights and learning rate). Optimization algorithms or strategies are responsible for reducing losses and providing the most accurate results. There are many types of optimizers.[84]

- ***Batch size***

Batch refers to the quantity being produced or planned to be produced. It is the total number of training examples present in a single batch.[85]

- ***Epochs***

In terms of artificial neural networks, an epoch refers to a cycle of the entire training data set. An epoch is often mixed with an iteration. The number of iterations is the number of batches or steps required to complete a partitioned data packet of training data. Epoch parameters refer to how many times the models aim to use the complete dataset for training.[86]

- ***Batch normalization***

Batch normalization is one of the most exciting latest innovations in optimizing deep neural networks. Batch normalization allows each layer of the network to learn independently of other layers. Because it has a slight regularization effect, it reduces overfitting.[87]

4- Discussion

We proposed a sentiment analysis model based on the ANN model Its purpose is to analyze sentiment more effectively than Gaussian-NB Model. This model proved to be highly effective and accurate in the analysis of feelings.

5- Application interface

Our interface is composed of four pages: Login page, Main page, Analyse Dataset page, and Analyse Sentence page.

Log-in page: to log in to the Smart Sentiment Analysis.

Main page: it presents the used tools: the dataset, the ANN model, and the anaconda environment, and you can select two choices to analyze the dataset or sentence.

Analyse Dataset page: to analyze reviews dataset.

Analyse Sentence page: to analyse santence review .

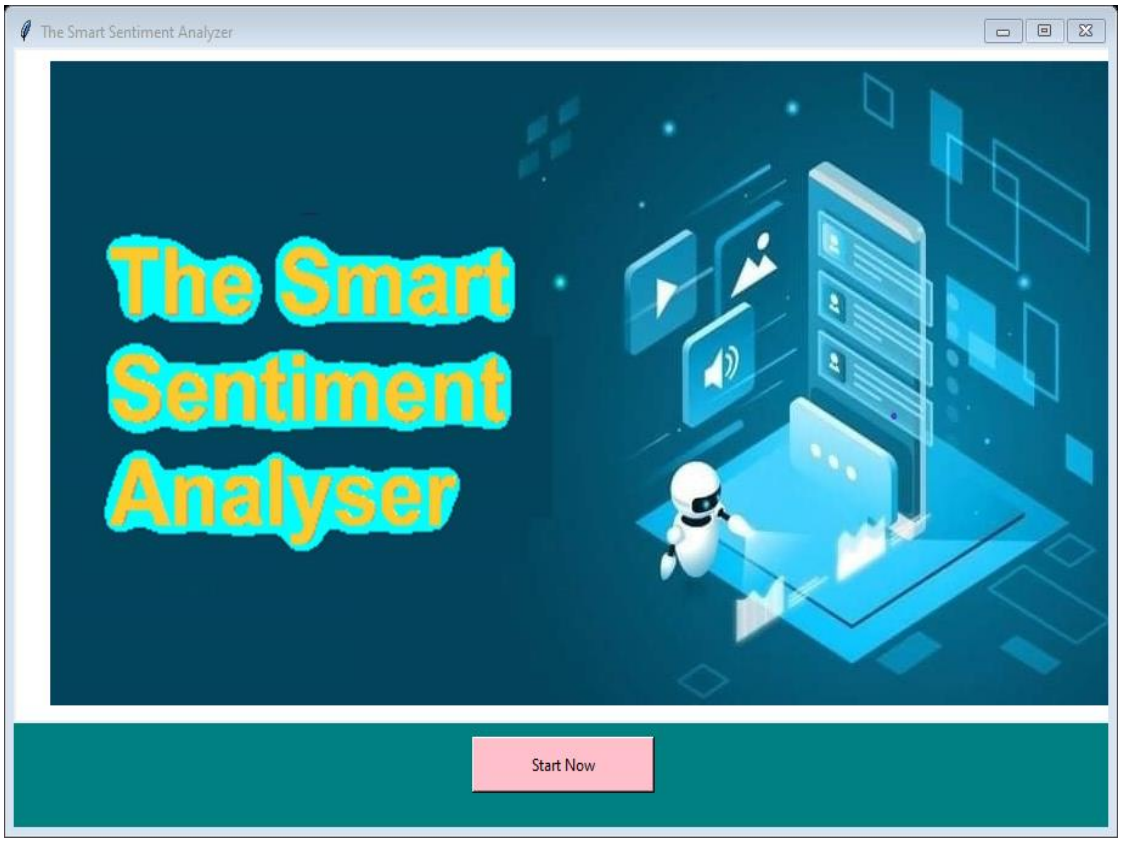


Fig 4.8 Log in page

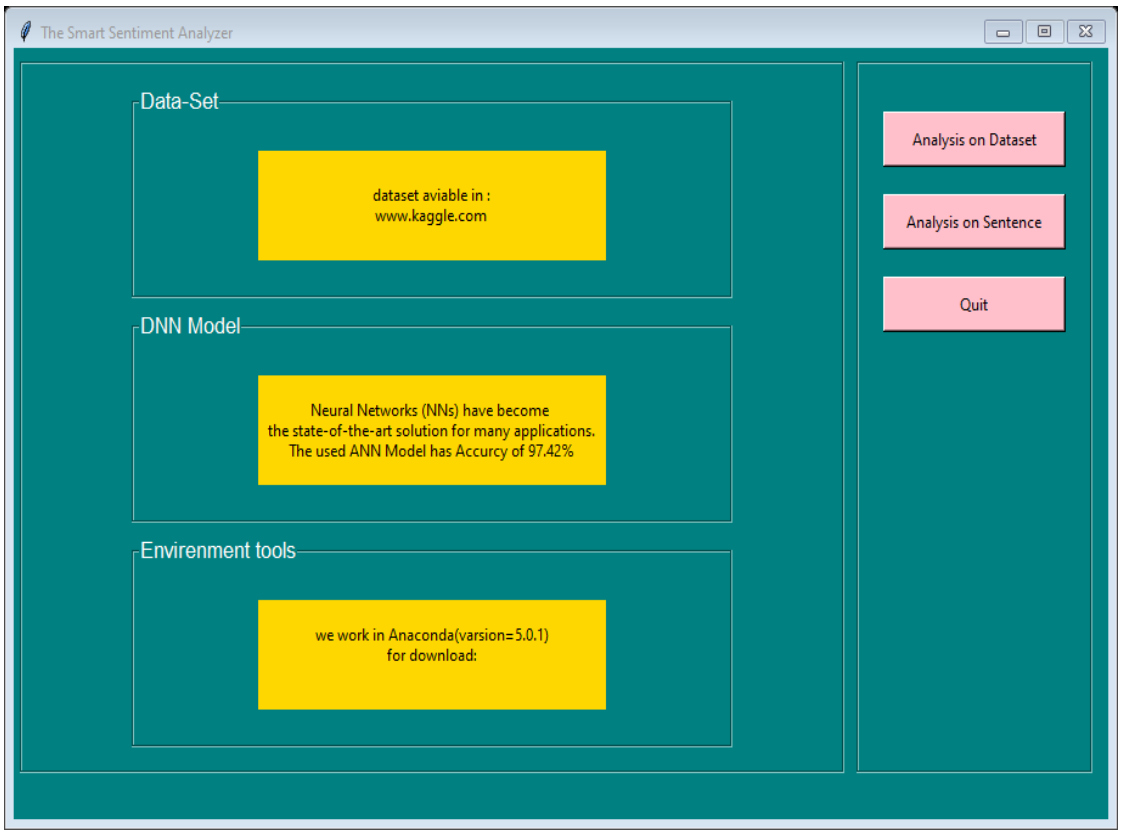


Fig 4.9 The main page

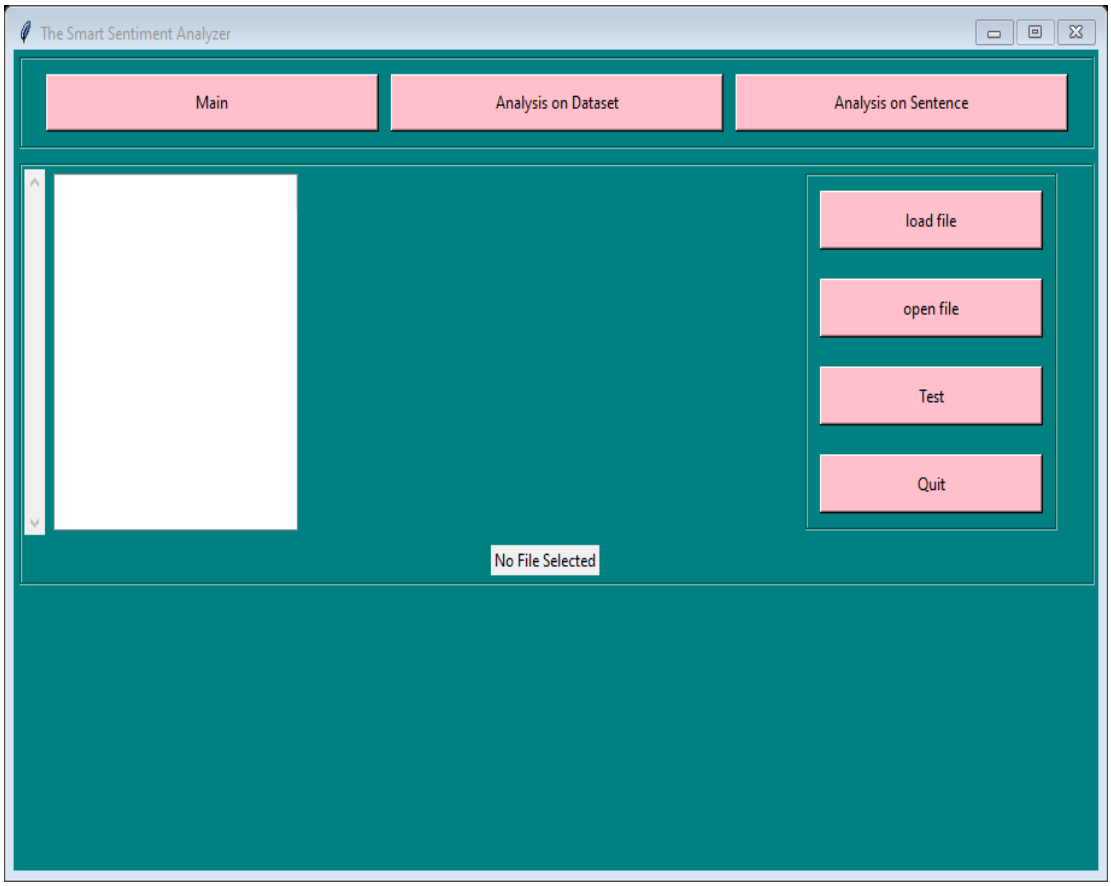


Fig 4.10 Analyse Dataset page

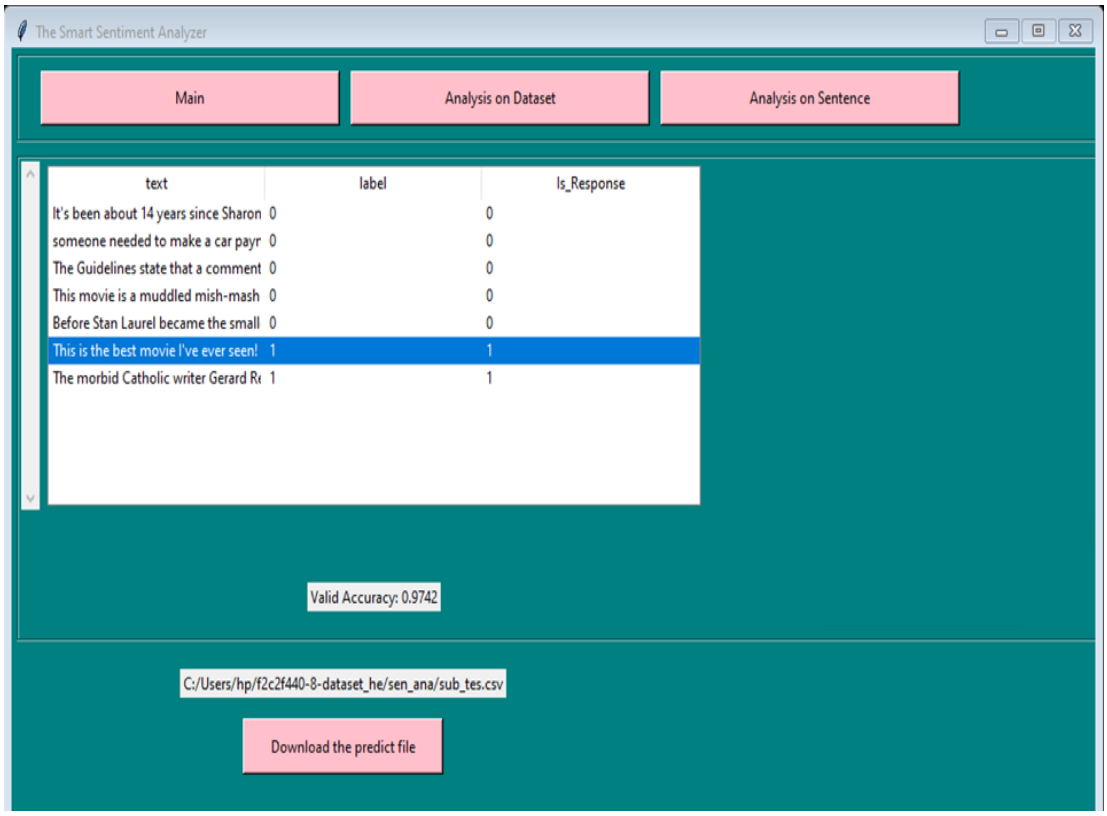


Fig 4.11 Analyse Dataset page

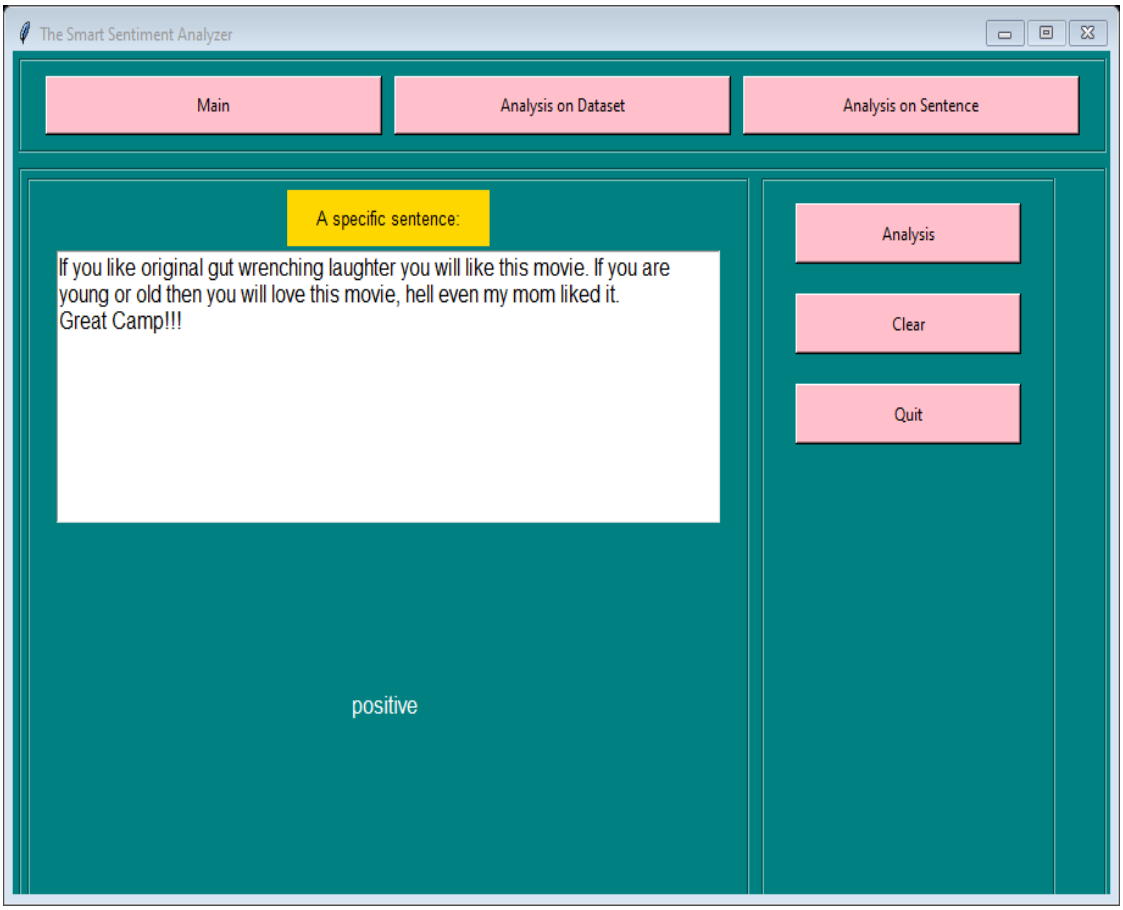


Fig 4.12 Analyse Sentence page

6- Conclusion

Our work is separated into two parts: the developed ANN model and our application interface. which is summarized as follows: We also specified the suggested model, the estimate of the suggested ANN model, and the interpretation of the obtained findings while describing the suggested model. In the last section, we showed basic concepts and our application interface.

GENERAL CONCLUSION

General Conclusion

Due to the huge number of real-world applications where uncovering people's opinions is vital in improved decision-making, the discipline of sentiment analysis is an intriguing new study topic. One of the most important aspects of this field is the development of methods for document-level sentiment analysis.

In this work, we proposed a model to analyze a dataset of movie reviews. We also presented sentiment deep learning to apply a ANN model of the movie reviews located in two different datasets. Our experimental approaches studied the accuracy of sentiment analysis.

This work in its current state faces more difficulties: first, the data set obtained for the film. Because the majority of related work worked on texts reviews.

The obtained results when applying the proposed ANN model on the training set and the test set are very high. Two performance measures are considered in this case, the loss value and the accuracy value. Our results are: 97.42% for training accuracy, 97.42% for validation accuracy, and the loss: 0.0026 for training set and 0.1157 for the validation set. It is clear that the loss value is very low against the accuracy which is very high and depends on the size of the used set. It is the raison for which the accuracy of the training set is higher than the accuracy of test set

For future work, we would like to extend this study to use other datasets such as Amazon dataset or tweeter dataset and use different feature selection methods. Furthermore, we may apply sentiment classification algorithms to detect fake reviews using various tools such as Python and Statistical Analysis System (SAS), and State; then we will evaluate the performance of our work with some of these tools.

REFERENCES

References

[1] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, ‘‘Lexicon based methods for sentiment analysis,’’ Comput. Linguistics, vol. 37, no. 2, pp. 267–307, 2011.

[2] A. Jurek, M. D. Mulvenna, and Y. Bi, ‘‘Improved lexicon-based sentiment analysis for social media analytics,’’ Secur. Informat., vol. 4, no. 1, p. 9, 2015

[3] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim, and I. A. Khan, ‘‘Lexiconenhanced sentiment analysis framework using rule-based classification scheme,’’ PLoS ONE, vol. 12, no. 2, Feb. 2017, Art. no. e0171649.

[4] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, and S. Massie, ‘‘Lexicon based feature extraction for emotion text classification,’’ Pattern Recognit. Lett., vol. 93, pp. 133–142, Jul. 2017.

[5] C. Dhaoui, C. M. Webster, and L. P. Tan, ‘‘Social media sentiment analysis: Lexicon versus machine learning,’’ J. Consum. Marketing, vol. 34, no. 6, pp. 480–488, Sep. 2017.

[6] C. S. Khoo and S. B. Johnkhan, ‘‘Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons,’’ J. Inf. Sci., vol. 44, no. 4, pp. 491–511, Aug. 2018.

[7] S. Zhang, Z. Wei, Y. Wang, and T. Liao, ‘‘Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary,’’ Future Gener. Comput. Syst., vol. 81, pp. 395–403, Apr. 2018.

[8] H. Keshavarz and M. S. Abadeh, ‘‘ALGA: Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs,’’ Knowl.-Based Syst., vol. 122, pp. 1–16, Apr. 2017.

[9] S. Feng, K. Song, D. Wang, and G. Yu, ‘‘A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs,’’ World Wide Web, vol. 18, no. 4, pp. 949–967, Jul. 2015.

[10] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. Venugopal, ‘‘Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier,’’ World Wide Web, vol. 20, no. 2, pp. 135–154, Mar. 2017.

[11] Z. Hai, G. Cong, K. Chang, P. Cheng, and C. Miao, ‘‘Analyzing sentiments in one go: A supervised joint topic modeling approach,’’ IEEE Trans. Knowl. Data Eng., vol. 29, no. 6, pp. 1172–1185, Jun. 2017.

[12] J. Singh, G. Singh, and R. Singh, ‘‘Optimization of sentiment analysis using machine learning classifiers,’’ Hum.-Centric Comput. Inf. Sci., vol. 7, no. 1, p. 32, 2017.

[13] F. Huang, S. Zhang, J. Zhang, and G. Yu, ‘‘Multimodal learning for topic sentiment analysis in microblogging,’’ Neurocomputing, vol. 253, pp. 144–153, Aug. 2017.

[14] M. R. Huq, A. Ali, and A. Rahman, ‘‘Sentiment analysis on Twitter data using KNN and SVM,’’ Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 6, pp. 19–25, 2017.

[15] W. Long, Y.-R. Tang, and Y.-J. Tian, ‘‘Investor sentiment identification based on the universum SVM,’’ Neural Comput. Appl., vol. 30, no. 2, pp. 661–670, Jul. 2018.

[16] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, ‘‘Deep convolution neural networks for Twitter sentiment analysis,’’ IEEE Access, vol. 6, pp. 23253–23260, 2018.

[17] D. Hyun, C. Park, M.-C. Yang, I. Song, J.-T. Lee, and H. Yu, ‘‘Target-aware convolutional neural network for target-level sentiment analysis,’’ Inf. Sci., vol. 491, pp. 166–178, Jul. 2019.

[18] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, ‘‘Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis,’’ Cogn. Comput., vol. 10, no. 4, pp. 639–650, Aug. 2018.

[19] H. Chen, S. Li, P. Wu, N. Yi, S. Li, and X. Huang, ‘‘Fine-grained sentiment analysis of Chinese reviews using LSTM network,’’ J. Eng. Sci. Technol. Rev., vol. 11, no. 1, pp. 174–179, 2018.

[20] S. Wen, H. Wei, Y. Yang, Z. Guo, Z. Zeng, T. Huang, and Y. Chen, ‘‘Memristive LSTM network for sentiment analysis,’’ IEEE Trans. Syst., Man, Cybern. Syst., to be published.

[21] F. Abid, M. Alam, M. Yasir, and C. Li, ‘‘Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter,’’ Future Gener. Comput. Syst., vol. 95, pp. 292–308, Jun. 2019.

[22] T. Chen, R. Xu, Y. He, and X. Wang, ‘‘Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,’’ Expert Syst. Appl., vol. 72, pp. 221–230, Apr. 2017.

[23] F. Hu, L. Li, Z.-L. Zhang, J.-Y. Wang, and X.-F. Xu, ‘‘Emphasizing essential words for sentiment classification based on recurrent neural networks,’’ J. Comput. Sci. Technol., vol. 32, no. 4, pp. 785–795, Jul. 2017.

[24] YING LI , JIN WANG, AND R. SIMON SHERRATT, Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning, Sentiment Analysis for E-Commerce Product Reviews Based on Sentiment Lexicon and Deep Learning, IEEE ACCESS, VOLUME 8, 2020, pp. 23522-23530,

[25] “Sentiment Analysis: Types, Tools, and Use Cases”, (21Sep, 2018) [Online], url: https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/

[26] Said G., Multilingual Text Categorization: Based on Machine Learning Algorithms and Ontologies,Noor Publishing, ISBN: 978-620-2-34305-3, Copyright © 2017 International Book Market Service Ltd., Member of OmniScriptum Publishing Group

[27] ” DEFINITIONS”,[Online], url : https://www.dictionary.com/browse/analysis

[28] ZIANI Amel, La recommandation via l’analyse d’opinions, Thèse Doctorat , Université de Badji Mokhtar Annaba,algeria,2018

[29] Akrouf, Samir, et al. "Social Network Analysis and Information propagation: A case study using Flickr and YouTube networks." International Journal of Future Computer and Communication 2.3 (2013): 246-252.

[30] ”Sentiment Analysis: A Definitive Guide”,[Online], url:https://monkeylearn.com/sentiment-analysis/

[31] ” LingmeiZhao.github.io”,[Online], url : https://lingmeizhao.github.io/blogs/Sentiment-Analysis-on-Hotel-Reviews.html

[32] “6 Sentiment Analysis Real-World Use Cases”, [Online], url: https://www.repustate.com/blog/sentiment-analysis-real-world-examples/

[33] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. Comput Linguist J 267–307

[34] Tong RM (2001) An operational system for detecting and tracking opinions in on-line discussions. In: Working Notes of the SIGIR Workshop on Operational Text Classification, pp 1–6

[35] Turney P, Littman M (2003) Measuring praise and criticism: inference of semantic orientation from association. ACM Transact Inform Syst J 21(4):315–346

[36] Said G., Abdelouahab M.,: Arabic Texts Categorization: Features Selection Basedon the Extraction of Words’ Roots, 5th IFIP International Conference on ComputerScience and its Applications (CIIA’2015), 20, 21 May 2015, LNCS SpringerConference, Tahar Moulay University, Saida - Algeria

[37] ” Natural Language Processing - Semantic Analysis”,[Online],url: https://www.tutorialspoint.com/natural_language_processing/natural_language_processing_semantic_analysis.htm

[38] Said G., Abdelouahab M.: Contextual Categorization of Documents Using a New Panoply of Similarity Metrics, International Conference on Advanced Technology &Sciences (ICAT’14), 12-15 August, 2014, Antalya, Turkey.

[39] ” machine learning”, [Online], url: https://searchenterpriseai.techtarget.com/definition/machine-learning-ML

[40] Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2020). A clarification of misconceptions, myths and desired status of artificial intelligence. arXiv:2008.05607.

[41] Ford, M. (2018). Architects of Intelligence: The truth about AI from the people building it. Packt Publishing Ltd.

[42] Àlex Asenjo Carvajal, Time to land prediction based on machine learning models, 17 de juliol de 2020

[43] ” Decision Trees”, [Online],url: https://scikit-learn.org/stable/modules/tree.html

[44] ” Decision Tree Introduction with example”, [Online], url: https://www.geeksforgeeks.org/decision-tree-introduction-example/

[45] ” Logistic Regression (LR) - PRIMO.ai”, Primo.ai,(2020). [Online]. Available: http://primo.ai/index.php?title=Logistic_Regression_(LR).

[46] 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 these doc.

[47] Brik, Youcef, et al. "Sequential feature selection with machine learning techniques for heart disease diagnosing." 9th (Online) International Conference on Applied Analysis and Mathematical Modeling (ICAAMM21) June 11-13, 2021, Istanbul-Turkey.

[48] Ladjal, Mohamed, and Mohamed Imed Khelil. "Application of machine learning techniques for predicting the WQI for water quality monitoring: a case study in Algeria." 9th (Online) International Conference on Applied Analysis and Mathematical Modeling (ICAAMM21) June 11-13, 2021, Istanbul-Turkey.

[49] "Linear SVM vs Nonlinear SVM high dimensional data ", [Online], url: https://stackoverflow.com/questions/44606126/linear-svm-vs-nonlinear-svm-high-dimensional-data

[50] "Yulia Gavrilova (February 17th, 2021) Naive Bayes Algorithm for Beginners", [Online],url:https://serokell.io/blog/naive-bayes-classifiers

[51] "Naive Bayes Classification using Scikit-learn", (2020). [Online]. url: https://www.datacamp.com/community/tutorials/naive-bayes-scikit-learn

[52] " K-Nearest Neighbor (K-NN)". [Online]. url: https://www.techopedia.com/definition/32066/k-nearest-neighbor-k-nn

[53] " Introduction to k-Nearest Neighbors (kNN) Algorithm". [Online]. url: https://ai.plainenglish.io/introduction-to-k-nearest-neighbors-knn-algorithm-e8617a448fa8

[54] "14 Different Types of Learning in Machine LearningNovember", (11, 2019) . [Online]. url: https://machinelearningmastery.com/types-of-learning-in-machine-learning/

[55] Page 695, Artificial Intelligence: A Modern Approach, 3rd edition, 2015.

[56] "ResearchGate", (2020). (PDF) Dive into Machine Learning. URL: https://www.researchgate.net/publication/328614973_Dive_Into_Machine_Learning

[57] Khelil, Mohamed Imed, and Mohamed Ladjal. "Hybrid Predictive Models for Water Quality Assessment Based on Water Quality Index Using ANN, LSSVM and multivariate statistical Methods." 9th (Online) International Conference on Applied Analysis and Mathematical Modeling (ICAAMM21) June 11-13, 2021, Istanbul-Turkey.

[58] A Deep Dive Into Deep Learning Techniques for Solving Spoken Language Identification Problems Himanish Shekhar Das, Pinki Roy, in Intelligent Speech Signal Processing, 2019

[59] Said G. Diabetic Patient Classification: An Efficient Algorithm Based on Deep Learning Approach, International Conference on Science, Engineering&Technology (ICSET) held in Istanbul, Turkey, 20-21 December, 2019 (Winner of Best Conference Paper)

[60] " Activation Functions in Neural Networks",(Sep 6, 2017), [Online]. Available: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

[61] ADOUANE Nour El Houda, Traffic signs recognition for autonomous cars, Master thesis, University of M'sila , 2019 /2020

[62] " How to Visualize a Neural Network in Python using Graphviz ?" https://www.geeksforgeeks.org/how-to-visualize-a-neural-network-in-python-using-graphviz/

[63] " Hidden Layer", [Online],url: https://deepai.org/machine-learning-glossary-and-terms/hidden-layer-machine-learning

[64] Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., & De, D. (2020). Fundamental Concepts of Convolutional Neural Network (pp. 519–567).UR.L: https://doi.org/10.1007/978-3-030-32644-9_36

[65] M. Peng, C. Wang, T. Chen, G. Liu and X. Fu, "Dual Temporal Scale Convolutional Neural Network for Micro-Expression Recognition", Frontiers in Psychology, vol. 8, 2017.

[66] Said G. Efficient Arabic handwritten character recognition based on machine learning and deep learning approaches, The 2 nd International Conference on Advances in Intelligent

[67] "Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network",(DEC 7, 2020), [Online],url: https://www.upgrad.com/blog/basic-cnn-architecture/

[68]"Recurrent neural networks deep dive", IBM Developer, 2020. [Online]. URL: https://developer.ibm.com/articles/cc-cognitive-recurrent-neural-networks/.

[69] "Understanding deep Convolutional Neural Networks with a practical use-case in "Tensorflow and Keras https://www.ahmedbesbes.com/blog/introduction-to-cnns

[70] "What is Python? Executive Summary", [Online]. URL: https://www.python.org/doc/essays/blurb/

[71] Sarkar, D., Bali, R. and Sharma, T., n.d. Practical Machine Learning with Python, p69.

[62] In: Researchgate. 2018. PYTHON CON ANACONDA. Bolivia: Aldo Ramiro Valdez Alvarado, pp.6, 7.

[73] https://repo.anaconda.com/archive/.

[74] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[75] https://www.scipy.org/

[76] what-is-pandas-in-python. URL https://www.educative.io/edpresso/what-is-pandas-in-python.

[77] What is NumPy, NumPy documentation. URL: https://numpy.org/doc/stable/user/whatisnumpy.html.

[78] Tosi, S., 2009. Matplotlib For Python Developers. Birmingham, U.K.: Packt Pub., p.8.

[79] TensorFlow, API Documentation. URL: https://www.tensorflow.org/api_docs.

[80] About keras. URL: https://keras.io/about/.

[81] https://docs.python.org/3/library/json.html

[82] What Is a CSV File, and How Do I Open It?",URL: https://www.howtogeek.com/348960/what-is-a-csv-file-and-how-do-i-open-it/

[83] Theanets.readthedocs.io. 2020. Loss Functions — Theanets 0.7.3 Documentation. URL: https://theanets.readthedocs.io/en/stable/api/losses.html

[84] Medium. 2020. Various Optimization Algorithms for Training Neural Network. URL: https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6.

[85] Swamidass, P., 2006. Encyclopedia of Production and Manufacturing Management. [New York?]: Springer, Boston, MA.

[86] DeepAI. 2020. Epoch. URL: https://deepai.org/machine-learning-glossary-and-terms/epoch.

[87] Ioffe, Sergey & Szegedy, Christian. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift

الملخص: تعتبر مراجعة منتجات العملاء حاسمة في تحديد ما إذا كان العميل سيشتري منتجًا أو يستخدم خدمة أم لا. تؤثر مراجعات الإنترنت والمدونات ومنصات الشبكات الاجتماعية الأخرى على خيارات العملاء وآرائهم. تُستخدم أنظمة التوصية الآن بشكل متكرر في العديد من مواقع التجارة الإلكترونية التجارية لمساعدة المستهلكين في التعامل مع مشكلة الحمل الزائد للمعلومات. يمكن للمستخدمين الحصول على توصيات مخصصة من خلال أنظمة التوصية، والتي يمكن أن تساعدهم في تحديد أفضل المنتجات التي يشترونها مع عدد كبير من الخيارات. في عملنا، قمنا بتطوير نموذج تحليل المشاعر على أساس نموذج ANN. والغرض منه هو تحليل استقطاب المشاعر لتعليقات الأفلام. ثبت أن هذا النموذج فعال للغاية ودقيق في تحليل المشاعر.

الكلمات المفتاحية: التعلم العميق ، تحليل المشاعر ، الشبكات العصبية، التعلم الآلي .

.

Abstract: Customer product reviews are critical in determining whether or not a customer will buy a product or utilize a service. Others' internet reviews, blogs, and social networking platforms influence customer choices and opinions. Recommender systems are now frequently used in many commercial sites to assist consumers in dealing with the problem of information overload. Users can get tailored recommendations via recommender systems, which can help them, make better selections on which product to buy from a large number of options. In our work, we developed a sentiment analysis model based on the ANN model. Its purpose is to analyze the sentiment polarity of movies comments. This model proved to be highly effective and accurate in the analysis of feelings.

Keywords: deep learning, Sentiment analysis, Neural Networks, Machine learning.

Résumé : Les avis des clients relatifs à des produits sont essentiels pour déterminer si un client achètera ou non un produit ou utilisera un service. Les avis d'autres internautes, les blogs et les plateformes de réseaux sociaux influencent sur les choix et les opinions des clients. Les systèmes de recommandation sont maintenant fréquemment utilisés dans de nombreux sites commerciaux du commerce électronique pour aider les consommateurs à faire face au problème de la surcharge d'informations. Les utilisateurs peuvent obtenir des recommandations personnalisées via les systèmes de recommandation, ce qui peut les aider à faire de meilleures sélections sur le produit à acheter parmi un grand nombre d'options. Dans notre travail, nous avons développé un modèle d'analyse des sentiments basé sur le modèle ANN. Son objectif est d'analyser la polarité des sentiments à partir des commentaires sur des films. Ce modèle s'est avéré être très efficace et précis dans l'analyse des sentiments.

Mots clés : apprentissage profond, analyse du sentiment, Réseaux de neurone, Apprentissage automatique.