**Arrow

Description automatically generated**

**Benha University**

**Faculty of Computers & Artificial intellengce**

**Social Media Monitoring (Web-based NLP Platform)**

A senior project submitted in partial fulfillment of the requirements for the degree of Bachelor of Computers and Informatics

**Computer Science Department,**

***Project Team***

1. Mohamed Reda lutfe
2. Kaream Abdelbadia

***Under Supervision of***

**Dr. Mustafa Abdul-Salam**

**Dr. Ahmed El-Sawy**

Benha, July 2021

**DECLARATION**

We hereby certify that this material, which we now submit for assessment on the program of study leading to the award of Bachelor of Computers and Artificial intelligence in Computer Science is entirely our own work, that we have exercised reasonable care to ensure that the work is original, and does not to the best of our knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of our work.

**Signed:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Date:** July 13th, 2021

**ACKNOWLEDGMENT**

Firstly, the success of our project happens when we believe in ourselves, believe in

doing the right thing to help other people, and the help of many people around us.

Secondly, we would like to thank our supervisor, Dr. Ahmed Hassan for his

encouragement, patience, guidance, and support when our project starts until we

finished the project; thank you, Dr. Ahmed Hassan for helping us and for always

wishing us the best

**ABSTRACT**

This project is an application to social media monitoring by scrapping data from sites

or files , The first part the user chose which way to track specific product by upload

file or track hash tag, The second is scrapping data about specific product and analyze

it by using natural language processing models, The third display the structured file

of data about specific product by two ways dashboard or structured file only.

Table of Contents

[**LIST OF FIGERS** 7](#_Toc76499158)

[**LIST OF TABELS** 8](#_Toc76499159)

[**LIST OF ACRONYMS/ABBREVIATIONS** 9](#_Toc76499160)

[**1** **INTRODUCTION** 10](#_Toc76499161)

[1.1 Importance of LSTM model 10](#_Toc76499162)

[1.2 PROBLEM STATEMENT 10](#_Toc76499163)

[1.3 OBJECTIVE 10](#_Toc76499164)

[1.4 MOTIVATION 11](#_Toc76499165)

[1.5 SCOPE 11](#_Toc76499166)

[1.6 STACKHOLDERS 11](#_Toc76499167)

[1.7 DELIVERABLES 12](#_Toc76499168)

[1.8 document organization 12](#_Toc76499169)

[**2** **Literature review** 13](#_Toc76499170)

[2.1 PREPROCESSING TECHNIQUES 13](#_Toc76499171)

[***2.1.1*** **PREPROCESSING TECHNIQUES** 13](#_Toc76499172)

[2.2 LSTM model 13](#_Toc76499173)

[***2.2.1*** **Recurrent Neural Networks** 13](#_Toc76499174)

[**3** **Methodology** 14](#_Toc76499175)

[3.1 Subtitle 1 (Heading 2) 14](#_Toc76499176)

[*3.2* Subtitle 2 (Heading 2) 14](#_Toc76499177)

[3.3 Equations 14](#_Toc76499178)

[**4** **System analysis and design** 16](#_Toc76499179)

[4.1 Assumptions 16](#_Toc76499180)

[4.2 SYSTEM REQUIREMENTS 16](#_Toc76499181)

[4.3 Use case diagram 16](#_Toc76499182)

[4.4 Data flow diagram 16](#_Toc76499183)

[4.5 Sequence diagram 16](#_Toc76499184)

[4.6 Activity diagram 16](#_Toc76499185)

[**5** **Web APPLICATION** 17](#_Toc76499186)

[**6** **IMPELEMINTATION OF COMPONANT** 18](#_Toc76499187)

[*6.1* *Equations* 18](#_Toc76499188)

[*6.2* *Equations* 18](#_Toc76499189)

[*6.3* *Equations* 18](#_Toc76499190)

[**7** **Conclusion** 19](#_Toc76499191)

[7.1 Future work 19](#_Toc76499192)

[**8** **REFRANCE** 20](#_Toc76499193)

# **LIST OF FIGERS**

Figure ‎1‑1 idea 9

Figure ‎2‑1 Recurrent Neural Network 11

Figure ‎2‑2Recurrent Neural Networks have loops 14

Figure ‎2‑3 The Problem of Long-Term Dependencies 15

Figure ‎2‑4 Problem of Long-Term Dependencies 15

Figure ‎2‑5The repeating module in a standard RNN contains a single layer. 16

Figure ‎2‑6The repeating module in an LSTM contains four interacting layers 16

# **LIST OF TABELS**

# **LIST OF** **ACRONYMS/ABBREVIATIONS**

Chapter ONE

# **INTRODUCTION**

## Importance of LSTM model

## PROBLEM STATEMENT

We will use natural language processing (NLP) and machine learning to transform mountains of hashtags, slang, and poor grammar into structured data and useful insights. Data analysts upload, process and analyze mountains of social text data in our platform to understand the conversations surrounding products, brands, people and services. Technology companies integrate our NLP APIs into their social listening product to deliver better insights to their own customers.

The problem that this project is trying to solve is that it makes it difficult to collect data for a particular product, whether from one site or from such a site, because it was applied manually. A group of people whose job is to track this product and so on, but in this project this problem will be solved and that The website, you can upload it to a file or do a hash tracking, and it will show you the organized data that you can benefit from in a more easy way.

## OBJECTIVE

The goal of this project was to develop a web page with NLP Models that helps companies (organizations) or people in tracking a specific product by scraping and compiling data and classifying it into negative and positive, and this saves on these institutions making people work to collect and analyze data

Website Features:

1- Easy to use

2- Save time

3. Effort saving

## MOTIVATION

The main motivation behind our work is that, until today, there’s no reliable system for automatically rating commenting of customers or users of specific product or specific service so our project is to help data analysts and technology firms throughout the world make social data more valuable. Data Analyst teams are using our platform to tell consumers' talks and feelings about products, brands, individuals, trends and services.

## SCOPE

This document covers the details of monitoring social media comments and rating it into negative and positive and number of spelling mistakes is out of scope . It also describes system with its components: A web-based information system.

## STACKHOLDERS

This project targets companies or organizations that are trying to track their products and know all opinions and classify them as positive and negative. This is what the project offers. Through this information, you can determine how to deal with these products and improve them, and so on, or know what is popular and through which companies start to see ways to improve or Providing products that match what is trendy. The project also targets people who want information so that they can determine the most appropriate product for them through the opinions that the project collects and categorizes from social media.

## DELIVERABLES

- Hardware devices: None

- Software:

1. A machine-learning model allow gradients to flow unchanged .
2. A web-based information system for user to upload data and analysis it . This system utilizes the LSTM machine learning model.

## document organization

This report will discuss all aspects of product tracking through social media monitoring, focusing on the data scrap aspect of websites and their analysis through NLP models. The first chapter provides a brief introduction to the project, problem statement, scope, outputs, and impact. In the second chapter, we present a general review of the most important points in the project. Chapter 3 describes. The fourth chapter explains Design and the components of the system. Chapter 4 . The sixth chapter presents the diagrams on which the project is based. Chapter 7 is the part about the web application and its components. Chapter 8 is an integration and general explanation of the components of the project. Finally, the conclusion and future work are presented in Chapter9

Chapter TWO

# **Literature review**

## PREPROCESSING TECHNIQUES

### **PREPROCESSING TECHNIQUES**

## LSTM model

### **Recurrent Neural Networks**

Chapter THREE

# **Methodology**

**We have here to explain step by step and how models work**

We

of thumb states that if it doesn’t sound good, then something must be wrong. When using MS Word you can change it to, “when there’s a red, green or blue wavy line, then something must be wrong.” Other proofreading pointers follow in the next subsections.

## Subtitle 1 (Heading 2)

Before using abbreviations or acronyms, make sure that the long name has been used first followed by the short name enclosed in parentheses. Afterwards, the abbreviations and acronyms can then be used alone. However, try not to use them when writing titles or subtitles.

Notice *Heading 2*’s format with a style of *Bold*, *Italics* and *Left Indent*. This format should be followed for all subtitles and even inner subtitles.

## Subtitle 2 (Heading 2)

Use metric system’s modern form, the International System of Units (SI), in writing units. Table 3.1 shows the seven SI base units. Don’t be mistaken for the acronym SI as it came from French, Le Système international d'unités.

Furthermore, always include zero before decimal points of numbers less than one such as “0.75” but not “.75”.

## Equations

Use the Insert Equation in MS Word 2010 when writing equations by clicking Insert tab and Equation icon beside the Symbol icon, as shown in Fig. 1.

circle while (2) is a quadratic equation. Notice that the word “equation” was used in the beginning of the statement when referring to (1) but was not included when mentioning a formula inside the statement like (2).

Chapter FOUR

# **System analysis and design**

## Assumptions

using abbreviations or acronyms, make sure that the long name has been used first followed by the short name enclosed in parentheses. Afterwards, the abbreviations and acronyms can then be used alone

## SYSTEM REQUIREMENTS

using abbreviations or acronyms, make sure that the long name has been used first followed by the short name enclosed in parentheses. Afterwards, the abbreviations and acronyms can then be used alone

## Use case diagram

## Data flow diagram

## Sequence diagram

## Activity diagram

Chapter FIVE

# **Web APPLICATION**

Chapter SIX

# **IMPELEMINTATION OF COMPONANT**

## *Equations*

using abbreviations or acronyms, make sure that the long name has been used first followed by the short name enclosed in parentheses. Afterwards, the abbreviations and acronyms can then be used alone

## *Equations*

using abbreviations or acronyms, make sure that the long name has been used first followed by the short name enclosed in parentheses. Afterwards, the abbreviations and acronyms can then be used alone

## *Equations*

using abbreviations or acronyms, make sure that the long name has been used first

followed by the short name enclosed in parentheses. Afterwards, the abbreviations and acronyms can then be used alone

Chapter SIVEN

# **Conclusion**

This project is an application to facilitate the tracking of products and their feedback through people or organizations. In designing our application, it does not require any costs and is easy to use, through some steps (Usage Guide):

1 - Open the website and register if you have an email and if not, create a new account

2- There are several options, either by uploading a file, or by following a hashtag or product

3- When choosing any method to track the product, the website makes a scrap of the product’s data and sends it to the analysis part

4- This part uses some of the nlp models to classify the positive and negative opinions that were scraped from the sites

5- The final result is displayed in the form of an organized file or in the form of a dashboard

6- The user's previous files and results can be displayed

## Future work

1 - The application supports many languages such as Arabic, French, Hindi……. etc

2- It will not be limited only to certain sites such as Instagram, Twitter and Amazon

3-make an mobile version of this project (android , ios)

Chapter EIGHT

# **REFRANCE**

|  |  |
| --- | --- |
| [1] | Brownlee, J., 2021. A Gentle Introduction to Long Short-Term Memory Networks by the Experts. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/> [Accessed 4 July 2021]. |
| [2] | (LSTM), I. and Saxena, S., 2021. *LSTM | Introduction to LSTM | Long Short Term Memory*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/> [Accessed 4 July 2021]. |
| [3] | En.wikipedia.org. 2021. Long short-term memory - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Long\_short-term\_memory> [Accessed 28 June 2021]. |
| [4] | Medium. 2021. LSTM: How To Train Neural Networks to Write like Lovecraft. [online] Available at: <https://towardsdatascience.com/lstm-how-to-train-neural-networks-to-write-like-lovecraft-e56e1165f514> [Accessed 4 July 2021]. |
| [5] | Y. Al-Lahham, K. Matarneh and M. Hassan, "Conditional Arabic Light Stemmer: CondLight", in The International Arab Conference on Information Technology, Yassmine Hammamet, Tunisia, 2017 |
| [6] | H. Rababah and A. T. Al-Taani, “An automated scoring approach for Arabic short answers essay questions,” in 2017 8th International Conference on Information Technology (ICIT), 2017. |
| [7] | A. Shehab, M. Faroun, and M. Rashad, "An Automatic Arabic Essay Grading System based on Text Similarity Algorithms", International Journal of Advanced Computer Science and Applications, vol. 9, no. 3, 2018 |
| [8] | A. A. Ewees, M. Eisa, and M. M. Refaat, "Comparison of cosine similarity and k-NN for automated essays scoring", IJARCCE, pp. 8669–8673, Dec. 2014. |
| [9] | M. Refaat, A. Ewees and M. Eisa, "Automated Assessment of Students' Arabic Free-Text Answers", International Journal of Intelligent Computing And Information Science, vol. 12, pp. 213-222, 2012. |
| [10] | W. H. Gomaa and A. A. Fahmy, "Arabic Short Answer Scoring with Effective Feedback for Students", International Journal of Computer Applications, vol. 86, no. 2, pp. 35–41, Jan. 2014. |
| [11] | G. Badaro et al., "EMA at SemEval-2018 Task 1: Emotion Mining for Arabic", in The 12th International Workshop on Semantic Evaluation, 2018. |
| [12] | N. Omar and R. Mezher, "A Hybrid Method of Syntactic Feature and Latent Semantic Analysis for Automatic Arabic Essay Scoring", Journal of Applied Sciences, vol. 16, no. 5, pp. 209–215, May 2016. |
| [13] | M. A. Zahran, A. Magooda, A. Y. Mahgoub, H. Raafat, M. Rashwan, and A. Atyia, "Word Representations in Vector Space and their Applications for Arabic", in Computational Linguistics and Intelligent Text Processing, Springer International Publishing, 2015, pp. 430–443 |
| [14] | S. Elkateb et al., "Building a WordNet for Arabic", in Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06), 2006. |
| [15] | C. Fellbaum, WordNet: An Electronic Lexical Database. 1998 |
| [16] | M. Boudchiche and A. Mazroui, “A hybrid approach for Arabic lemmatization,” International Journal of Speech Technology, Jul. 2018 |
| [17] | B. Hegde, N. H S and M. Prakash, "Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets", International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 6, pp. 1876-1880, 2018. |
| [18] | M. Alghamdi, M. Alkanhal, M. Al-Badrashiny, A. Al-Qabbany, A. Areshey, and A. Alharbi, "A hybrid automatic scoring system for Arabic essays", AI Communication (AIC), vol. 27, no. 2, pp. 103–111, 2014 |
| [19] | K. Taghva, R. Elkhoury, and J. Coombs, "Arabic stemming without a root dictionary", in International Conference on Information Technology: Coding and Computing (ITCC’05) - Volume II, 2005 |
| [20] | A. Magooda, M. Zahran, M. Rashwan, H. Raafat and M. Fayek, "Vector Based Techniques for Short Answer Grading", in Proceedings of the Twenty-Ninth International Florida Artificial Intelligence Research Society Conference, 2016. |
| [21] | "Stanford CoreNLP – Natural language software | Stanford CoreNLP", Stanfordnlp.github.io.[Online].Available:https://stanfordnlp.github.io/CoreNLP. [Accessed 4 July 2021] |
| [22] | Q. W. Bsoul and M. Mohd, "Effect of ISRI Stemming on Similarity Measure for Arabic Document Clustering", in Information Retrieval Technology, Springer Berlin Heidelberg, 2011, pp. 584–593 |
| [23] | H. Mubarak, "Build fast and accurate lemmatization for Arabic", arXiv preprint arXiv:1710.06700., 2017 |
| [24] | A. A. Freihat, M. Abbas, G. Bella, and F. Giunchiglia, "Towards an Optimal Solution to Lemmatization in Arabic", Procedia Computer Science, vol. 142, pp. 132–140, 2018 |
| [25] | M. Boudchiche and A. Mazroui, “A hybrid approach for Arabic lemmatization,” International Journal of Speech Technology, Jul. 2018. |
| [26] | A. Kadim and A. Lazrek, "Parallel HMM-Based Approach for Arabic Part of Speech Tagging", The International Arab Journal of Information Technology, vol. 15, no. 2, pp. 341-351, 2018 |
| [27] | R. Alharbi, W. Magdy, K. Darwish, A. AbdelAli and H. Mubarak, "Part-of-Speech Tagging for Arabic Gulf Dialect Using Bi-LSTM", in the Eleventh International Conference on Language Resources and Evaluation (LREC-2018), 2018, pp. 3925-3932 |
| [28] | Yousif and M. Al-Risi, "PART OF SPEECH TAGGER FOR ARABIC TEXT BASED SUPPORT VECTOR MACHINES: A REVIEW", ICTACT JOURNAL ON SOFT COMPUTING, vol. 9, no. 2, pp. 1867-1873, 2019 |
| [29] | Colah.github.io. 2021. Understanding LSTM Networks -- colah's blog. [online] Available at: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> [Accessed 5 July 2021]. |
| [30] | Medium. 2021. Neural Language Models. [online] Available at: <https://towardsdatascience.com/neural-language-models-32bec14d01dc> [Accessed 5 July 2021]. |
| [31] | Bioinf.jku.at.2021. [online] Available at: <https://www.bioinf.jku.at/publications/older/2604.pdf> [Accessed 5 July 2021]. |
| [32] | Sdv |
| [33] | Sdv |
| [34] | Sdvsd |
| [35] | Sadf |
| [36] | Sdfsad |
| [37] | Sdfsdf |
| [38] | Sdfs |
| [39] | Dsaf |
| [40] | Sdfas |
| [41] | Sdfa |
| [42] | Asdf |
| [43] | Asdfa |
| [44] | Sdaf |
| [45] | Asdf |
| [46] | Asdf |