

Arabic Sentiment Analysis (ASA) Using Deep Learning Approach

Abdulahkeem Q. Albayati Student of Master Computer Engineering Department University of Technology Baghdad, Iraq Email: 10466@uotechnology.edu.iq	Ahmed S. Al-Araji Prof. Dr. Computer Engineering Department University of Technology Baghdad, Iraq Email: 60166@uotechnology.edu.iq	Saman H. Ameen Lecturer Dr. Computer Engineering Department University of Technology Baghdad, Iraq Email: Saman7011@yahoo.com
--	--	--

ABSTRACT

Sentiment analysis is one of the major fields in natural language processing whose main task is to extract sentiments, opinions, attitudes, and emotions from a subjective text. And for its importance in decision making and in people's trust with reviews on web sites, there are many academic researches to address sentiment analysis problems. Deep Learning (DL) is a powerful Machine Learning (ML) technique that has emerged with its ability of feature representation and differentiating data, leading to state-of-the-art prediction results. In recent years, DL has been widely used in sentiment analysis, however, there is scarce in its implementation in the Arabic language field. Most of the previous researches address other languages like English. The proposed model tackles Arabic Sentiment Analysis (ASA) by using a DL approach. ASA is a challenging field where Arabic language has a rich morphological structure more than other languages. In this work, Long Short-Term Memory (LSTM) as a deep neural network has been used for training the model combined with word embedding as a first hidden layer for features extracting. The results show an accuracy of about 82% is achievable using DL method.

Keywords: Deep Learning (DL), Machine Learning (ML), Arabic Sentiment Analysis (ASA), word embedding, Long-Short Term Memory (LSTM), features.

تحليل المشاعر للغة العربية من خلال منهج التعلم العميق

عبد الحكيم قصي علي البياتي
طالب ماجستير
قسم هندسة الحاسوب-الجامعة التكنولوجية

احمد صباح الأعرجي
أستاذ دكتور
قسم هندسة الحاسوب-الجامعة التكنولوجية

سامان حميد أمين
مدرس دكتور
قسم هندسة الحاسوب-الجامعة التكنولوجية

الخلاصة

إن تحليل المشاعر هو أحد المجالات الرئيسية في مجال معالجة اللغة الطبيعية التي تتمثل مهمتها الرئيسية في استخراج المشاعر والآراء والمواقف والعواطف من نص شخصي. ولأهميته في صنع القرار وثقة الناس في إجراء المراجعات على مواقع الويب ، فهناك العديد من الأبحاث الأكاديمية لمعالجة مشكلات تحليل المشاعر.

*Corresponding author

Peer review under the responsibility of University of Baghdad.

<https://doi.org/10.31026/j.eng.2020.06.07>

2520-3339 © 2019 University of Baghdad. Production and hosting by Journal of Engineering.

This is an open access article under the CC BY4 license <http://creativecommons.org/licenses/by/4.0/>.

Article received: 16/12/2019

Article accepted: 19/1/2020

Article published:1/6 /2020



التعلم العميق عبارة عن تقنية ذات قدرة عالية من تعلم الآلة ظهرت لتمثيل الميزات وتمييز البيانات، مما أدى إلى أحدث النتائج في معالجة مسألة التنبؤ. ففي السنوات الأخيرة، تم استخدام التعلم العميق على نطاق واسع في تحليل المعنويات، ولكن هناك ندرة في تنفيذه في مجال اللغة العربية حيث تتناول معظم الأبحاث السابقة لغات أخرى مثل اللغة الإنجليزية. لذلك يتناول النموذج المقترح تحليل المشاعر العربية (ASA) باستخدام نهج التعلم العميق. فتحليل المعنويات العربية هو مجال مليء بالتحديات حيث أن اللغة العربية ذات بنية غنية أكثر من غيرها من اللغات. وفي هذا العمل، تم استخدام الذاكرة طويلة المدى (LSTM) كشبكة عصبية عميقة لتدريب النموذج مع دمج الكلمات (Word embedding) كطبقة أولية مخفية لاستخراج الميزات. أظهرت النتائج دقة حوالي 82% قابلة للتحقيق باستخدام نموذج التعلم العميق. الكلمات الرئيسية: التعلم العميق، تعلم الآلة، تحليل المشاعر العربية، الذاكرة طويلة المدى، دمج الكلمات، الميزات.

1. INTRODUCTION

Sentiment Analysis (SA), or sometimes called Opinion Mining (OM), is a field of Natural Language Processing (NLP) whose goal is to extract the emotion, sentiment or more general opinion expressed in a human-written text. The text mostly derives from social media, product reviews and blogs (Korovesis, 2018). Opinions and emotions play a central role in human life. They have the ability to describe the way the individuals think, behave, and act. SA has many trending applications in various real-life fields including finance, marketing, political science, health science, communications, and even history (Kaseb & Ahmed, 2016).

In general, SA can be carried out at three various scope levels (Kolkur, et al., 2015): level of document, level of sentence, and level of sub-sentence (aspect level). In this work, the level of sentence is used for ASA to determine a sentence polarity if it is **positive** or **negative**. Due to the importance of SA, numerous research studies have been conducted in this field. Yet, most of these performed researches have concentrated on English language and other Indo-European languages (Al-Sallab, et al., 2017). In fact, very few researches have involved the analysis of sentiment in rich morphological languages, such as Arabic. Arabic language has complex nature in addition to the lack of its resources and different dialects that give challenges to the advances in ASA research (Heikal, et al., 2018). Despite these challenges, increasing Arabic number of users for the internet and the exponential expansion of the Arabic online content are the things that heightened the attention of numerous researchers according to SA over the last decade (Boudad, et al., 2017).

Lately, the domain of DL has witnessed significant accomplishments in the branch of SA and is seen as the most advanced model in many languages (Heikal, et al., 2018) & (Ain, et al., 2017). Accordingly, the current work attempts to investigate one of the DL methods with the purpose of enhancing the accuracy of predicting the sentiment polarity of Arabic sentence. Moreover, the effect of different network parameters is investigated. The proposed model consists of multiple stages. First of all, the selected dataset is preprocessed to remove noises in it. Then a word embedding layer is used to convert words in the texts to sequences of numbers in order to prepare the data for the next layer. Next, LSTM is used to processing the data. Finally, the output from LSTM layer is fed to SoftMax layer to normalize the results and classify the sentiment of the input sentence to be either Positive or Negative.

The remainder of this paper organized as follows: in section 2, the related works are discussed in the SA field. In section 3, the system's main components and model architecture are illustrated. After that, the conducted experiments and their results are displayed, in section 4. Finally, a brief conclusion is given based on this work and the obtained results.

2. RELATED WORKS

This section presents an overview of different approaches proposed to perform SA in Arabic or in English.

The authors in (Shoukry & Rafea, 2012) are interested in sentiment classification in the Arabic language at the sentence level in which the aim is to classify a sentence as holding an overall

positive, negative or neutral sentiment with regards to the given target. used ML approach (Naïve Bayes (NB) & Support Vector Machine (SVM)).

In (Wang, et al., 2015), the authors used LSTM to predict polarity from tweets in twitter in English language by simulating the interactions of words during the compositional process. Multiplicative operations between word embeddings through gate structures are used to provide more flexibility and to produce better compositional results compared to the additive ones in simple Recurrent Neural Network (RNN).

In (El-Beltagy, et al., 2016), the authors presented a model for carrying out ASA by augmenting ML (using complement NB) approach with a set of features derived from an Arabic sentiment lexicon as well as from the text itself.

In (Al-Azani & El-Alfy, 2017) the authors investigated various DL models based on Convolutional Neural Network (CNN) and LSTM for SA of Arabic microblogs. They trained neural language models using two different word2vec based techniques: Continuous Bag of Word (CBOW) and skip-gram. In addition, the experiments showed that LSTM performs better than CNN.

The authors in (Heikal, et al., 2018) used an ensemble model, combining CNN and LSTM to predict the polarity of Arabic tweets on the Arabic Sentiment Tweets Dataset (ASTD). They achieved an F1-score of 64.46% and an accuracy of 65.05% on ASTD dataset.

3. SYSTEM ARCHITECTURE

In this work, there is an attempt to work with ASA a type of Arabic Natural Language Processing (ANLP) by DL model. The dataset used at this work is taken from the Internet and is called a Large-Scale Arabic Book Reviews (LABR) (Aly & Atiya, 2013), it contains book reviews in 16448 rows labeled as positive (1) or negative (0).

In this section, the main components of the system architecture are illustrated. First, the sentences of the dataset are entered into preprocessing to remove any noise in each raw of text that has an effect on the final output prediction of the model negatively. And then texts are converted to sequence numbers by one of word embedding methods for preparing the dataset to enter the selected deep neural network that is LSTM. Finally, the output from LSTM layer is fed to SoftMax layer to normalize the results and classify the sentiment of the input sentence to be either Positive or Negative, as shown in the Fig.1 below. Also, there should be attempts to change some parameters and hyperparameters in the model to see how they effect on results. In addition to that, the preprocessing will be omitted to see its effect on the model.

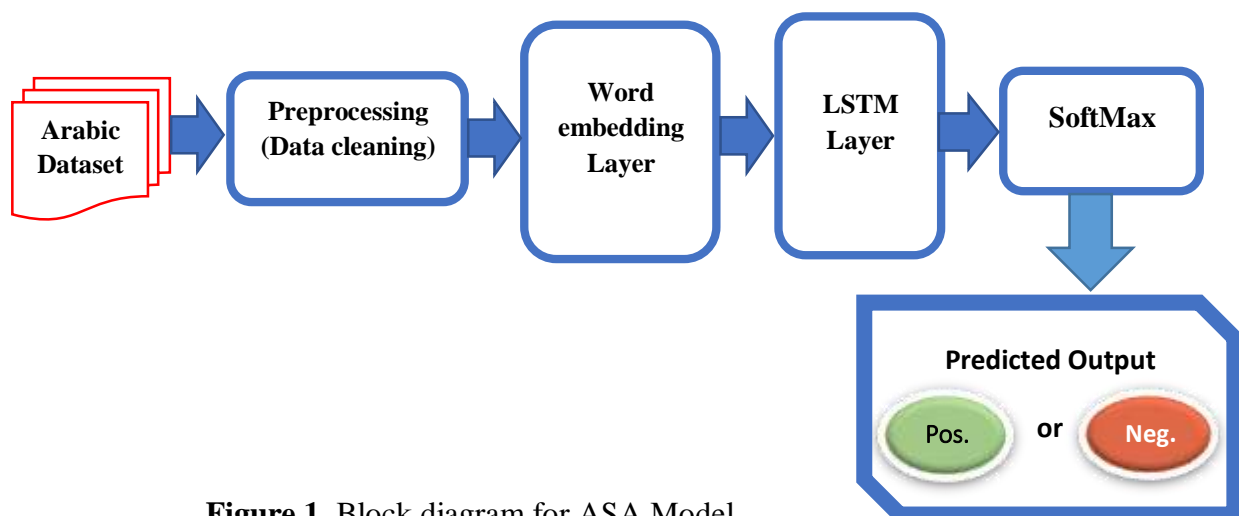


Figure 1. Block diagram for ASA Model



3.1 Text Preprocessing

Datasets are usually noisy and have many non-desirable characters and symbols that may have negative effects on learning quality, therefore it is essential to clean these datasets to make the process of learning by the neural network easier. Nevertheless, it should be clear that cleaning of datasets will remove some information from the dataset itself, which may be crucial for a specific application or analysis.

Data preparation and enhancement on the dataset in this work mainly consist of the following operations:

1. Removing any words not in Arabic script.
2. Orthographic Normalization
 - I. Remove diacritics.
 - II. Remove repetition (elongation).
 - III. Normalize some Arabic characters and punctuations.
3. Stemming the texts
4. Removing stopwords and re-join words.

3.2 Embedding Layer

This is the third block in **Fig.1** and it represents the first hidden layer in this model; the word embeddings method is used in this work.

In briefly, word embeddings give a way to use an efficient, dense representation in which similar words have a similar encoding. Importantly, there is no need to specify this encoding by hand. An embedding is a dense vector of floating-point values (the length of the vector parameter can be specified). Instead of specifying the values for the embedding manually, they are trainable parameters (weights learned by the model during training, in the same way a model learns weights for a dense layer). The resulting embedded vectors are representations of categories where similar categories (relative to the task) are closer to one another.

There are many strategies for learning word embeddings such as: Word2vec (by Google), Glove (by Stanford), Fastest (by Facebook) and Embedding layer (by Keras). One of them (embedding layer created by Keras library) has been applied in this work. It is the first hidden layer of the model to prepare data to be input value for the next layer of the model (LSTM layer).

3.3 LSTM Layer

Recently, there have been advances in DL concerning different fields such as NLP, computer visions, speech recognition and others. So many DL approaches have been used in SA domain to increase its accuracy and to give efficient results closer to reality (**Al-Ayyoub, et al., 2018**). There are different types of neural networks in DL like CNN, RNN, LSTM and Gated Recurrent Unit (GRU) that could be used for addressing ASA as a binary or a multi-class classification problem (**Ain, et al., 2017**).

LSTM is a special kind of RNN created by authors in (**Hochreiter & Schmidhuber, 1997**). In the biological intelligence, the information is analyzing incrementally but with keeping an internal model of what is being analyzed, developed based on past data and continually updated each time new data is considered. RNN works in the same manner but in a very basic mechanism (**Al-Araji,**

et al., 2011): that is, it deals with sequences as feed-forward neural network by iterating through the sequence elements and preserving a state that has data belongs to what it has lastly noticed (CHOLLET, 2018) & (Al-Araji, 2015). All RNNs have the configuration like a chain of repeating modules of neural network (Al-Araji, 2014). Such repeating module, in typical RNNs, have a quite basic structure, such as a single-tanh layer as shown in Fig. 2:

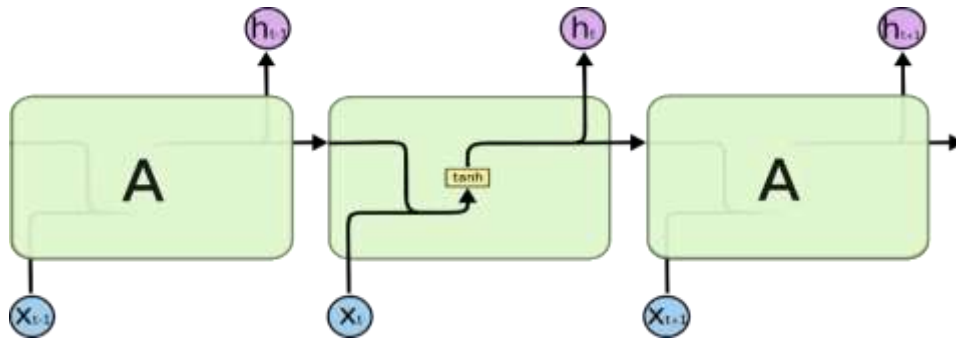


Figure 2. RNN (Goyal, et al., 2018).

However, RNNs suffer from the problem of vanishing gradients, which hampers learning of long data sequences. The gradients carry information used in the RNN parameter update. When the gradient becomes smaller and smaller, the parameter updates become insignificant which means no real learning is done. A type of RNN is LSTM that has capable to resolve the problem of RNN (Mueller & Massaron, 2019) and (Al-Araji & Dagher, 2015).

LSTM has the ability of learning long-term dependencies that is very useful in remembering sequences of texts in SA by multiple *gates* (Input gate, Forget gate, and Output gate) as shown in Fig. 3, which represent quite useful choice for controlling the information going through model (Dagher, 2018). It suggests a very elegant solution to the vanishing gradient problem. In this paper, this deep neural network is LSTM. It is used for addressing ASA by DL model.

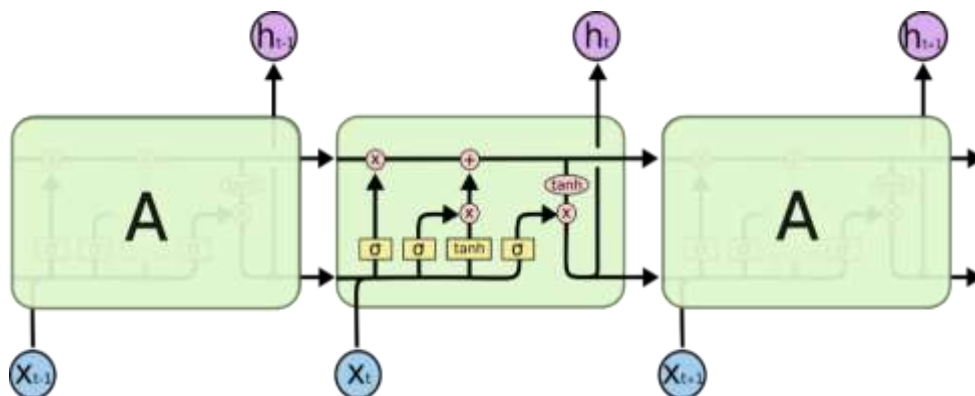


Figure 3. LSTM (Goyal, et al., 2018).

3.4 SoftMax

SoftMax has been used as activation function and it represents the output layer of the model. The main merit of the SoftMax activation function is its ability to program the network's output to describe the probability that the input will be into which class. Without the abilities of the SoftMax



function, the outputs of neuron are merely values at numeric scale where the winning class is determined by the largest value (Heaton, 2015). It takes the output of LSTM layer and normalize it from real numbers into probabilities between (0 and 1), in order to clarify each input value to the model for which class it belongs. SoftMax function has illustrated mathematically in Eq. (1) below (Heaton, 2015):

$$S_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{1}$$

Where i = the index of the output neuron being calculated,
 j = the indexes of all neurons in the group/level of layer,
 Z = designates the array of output neurons.

4. EXPERIMENTS & RESULTS

In this section, Experiments on ASA model and its results will be discussed, also how changing some parameters affect the accuracy of the results.

4.1 First Experiment

First experiment has been applied with the following properties:

- 1- Dataset: LABR (with 16448 samples divided to 67% for training, 16% for validation, and 17% for testing)
- 2- NN: LSTM
- 3- Input dim: 10000 words
- 4- Lstm_out: 100 vectors
- 5- Epochs: 10
- 6- Batch size: 256

Performance of the model trained has been shown in Fig. 4 below. The Accuracy for this experiment is 80% and the F-Score is 80.2%. The learning of the model on training data is increasing gradually after first epoch and it reached to 96% in the 10th epoch. While for the test data, the model needs more learning because its good only in training data but not showing good results on unseen data (Test data), as shown below in Fig. 5:

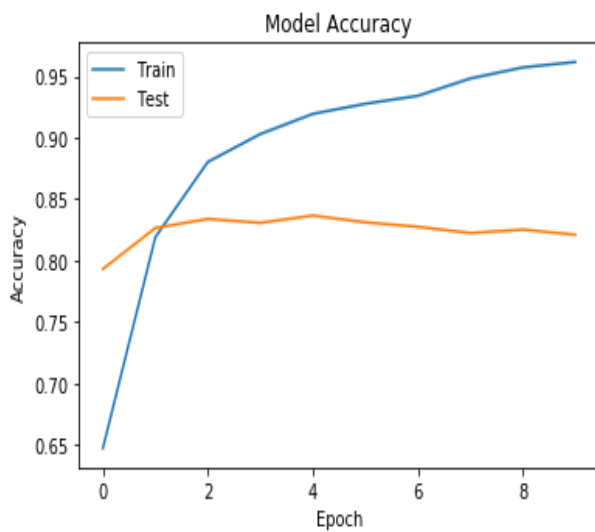


Figure 4. Model Accuracy for LABR dataset.

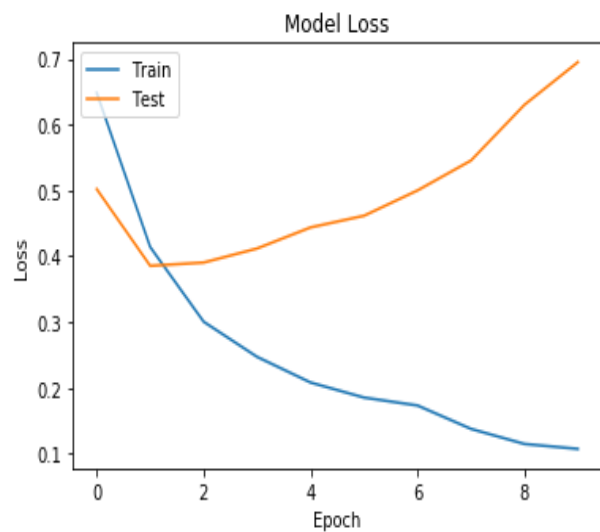


Figure 5. Model Loss for LABR dataset.



4.2 Second Experiment

This experiment is made on the same dataset but with different hyper-parameters like (Batch size) and (lstm-out):

4.2.1 Change of Batch size

To see the effect of changing one of hyper-parameters (Batch size) on the performance of the dataset in the **Table 1.** below:

Table 1. 2nd experiment with changing batch size.

No.	Batch size	Accuracy	F-Score
1	32	81%	81%
2	64	81%	80.9%
3	128	80%	80%
4	256	80%	80.2%

4.2.2 Change of lstm_out with fixed batch size

To see the effect of changing another hyper-parameter (lstm_out) on the performance of the dataset in the table (2) (Note: Batch size =256):

Table 2. 2nd experiment with changing lstm_out but fixed batch size.

No.	lstm_out	Accuracy	F-Score
1	50	82%	81.6%
2	75	81%	80.8%
3	100	80%	80.2%

5. CONCLUSION

In this work, one of the DL models has been presented to tackle ASA problem. LSTM neural networks are used with word embedding layer rather than hand-crafted features for ASA model. This model only relies on a pre-trained embedding layer representation, despite the complexity and richness morphology of Arabic language. Multiple experiments have been done by changing in hyper-parameters of the model and compared results between them to see effects on the model with different values. The best accuracy and F-score have been achieved when hyper-parameter lstm_out =50 and batch size = 256; in this case, the accuracy = 82% and F-score = 81.6%. There are many dimensions for this current work could be useful, such as: in politics, this program can be used to predict Arab opinions on current political events so that this field will be of importance in decision-making. Also in business, Arabic companies and others that sell their products to Arab consumers will benefit from such a work to automatically collect opinions about their products and services for developing them. As future enhancements on this work, some of the state-of-the-art neural networks like Transformers by (Vaswani, et al., 2017) and Bidirectional Encoder Representations from Transformers (Bert) by (Devlin, et al., 2018) can be used in this model instead of LSTM to try to improve the efficiency of the model, also there are



multiple methods for word embedding like: word2vec and glove can be used with this model see if there are any effects on the results.

REFERENCES

- Ain, Q. T. et al., 2017. Sentiment Analysis Using Deep Learning Techniques: A Review. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 8(6), pp. 424-433.
- Al-Araji, A. S., 2014. Applying Cognitive Methodology in Designing On-Line Auto-Tuning Robust PID Controller for the Real Heating System. *Journal of Engineering*, 20(9), pp. 43-61.
- Al-Araji, A. S., 2015. A Cognitive PID Neural Controller Design for Mobile Robot Based on Slice Genetic Algorithm. *Engineering & Technology Journal*, 33(1), pp. 208-222.
- Al-Araji, A. S., Abbod, M. F. & Al-Raweshidy, H. S., 2011. *Design of an adaptive nonlinear PID controller for nonholonomic mobile robot based on posture identifier*. s.l., IEEE International Conference on Control System, Computing and Engineering (ICCSCE), pp. 337-342.
- Al-Araji, A. S. & Dagher, K. E., 2015. Cognitive Neural Controller for Mobile Robot. *Iraqi Journal of Computer, Communication, Control and Systems Engineering*, 15(1), pp. 46-60.
- Al-Ayyoub, M. et al., 2018. Deep learning for Arabic NLP: A survey. *Journal of Computational Science*, May, Volume 26, pp. 522-531.
- Al-Azani, S. & El-Alfy, E.-S. M., 2017. *Hybrid deep learning for sentiment polarity determination of arabic microblogs*. Springer, Cham, s.n., pp. 491-500.
- Al-Sallab, A. et al., 2017. AROMA: A Recursive Deep Learning Model for Opinion Mining in Arabic as a Low Resource Language. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, September.16(4).
- Aly, M. & Atiya, A., 2013. LABR: A Large Scale Arabic Book Reviews Dataset. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, August, p. 494-498.
- Boudad, N., Faizi, R., Thami, R. O. H. & Chiheb, R., 2017. Sentiment analysis in Arabic: a review of the literature. *Ain Shams Engineering Journal*, 9(4), pp. 2479-2490.
- CHOLLET, F., 2018. *Deep Learning with Python*. Shelter Island in New York: Manning Publications Co..
- Dagher, K. E., 2018. A Nonlinear MIMO-PID Neural Controller Design for Vehicle Lateral Dynamics Model based on Modified Elman Neural Network. *Journal of Engineering*, 24(1), pp. 173-191.



- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K., 2018. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. s.l., arXiv preprint arXiv:1810.04805..
- El-Beltagy, S. R., Khalil, T., Halaby, A. & Hammad, M., 2016. *Combining Lexical Features and a Supervised Learning Approach for Arabic Sentiment Analysis*. Springer, Cham, s.n., pp. 307-319.
- Goyal, P., Pandey, S. & Jain, K., 2018. *Deep Learning for Natural Language Processing: Creating Neural Networks with Python*. s.l.:Apress.
- Heaton, J., 2015. *Artificial Intelligence for Humans, Volume 3: Neural Networks and Deep Learning*. 3rd ed. s.l.:Heaton Research, Inc..
- Heikal, M., Torki, M. & El-Makky, N., 2018. Sentiment analysis of Arabic Tweets using deep learning. *Procedia Computer Science*, 142(1), pp. 114-122.
- Hochreiter, S. & Schmidhuber, J., 1997. Long Short-Term Memory. *Neural computation*, November, 9(8), pp. 1735-1780.
- Kaseb, G. S. & Ahmed, M. F., 2016. Arabic Sentiment Analysis approaches: An analytical survey. *International Journal of Scientific & Engineering Research*, 7(10), pp. 1871-1874.
- Kolkur, S., Dantal, G. & Mahe, R., 2015. Study of different levels for sentiment analysis. *International Journal of Current Engineering and Technology*, 5(2), pp. 768-770.
- Korovesis, K., 2018. *Sentiment Analysis for Tweets*, Athens: Athens University.
- Mueller, J. P. & Massaron, L., 2019. *Deep Learning For Dummies*. Hoboken, New Jersey: John Wiley & Sons, Inc..
- Shoukry, A. & Rafea, A., 2012. *Sentence-level Arabic sentiment analysis*. Denver, CO, USA, IEEE, pp. 546-550.
- Vaswani, A. et al., 2017. *Attention is all you need*. Long Beach, CA, USA., Advances in neural information processing systems (NIPS 2017), pp. 5999-6009.
- Wang, X. et al., 2015. Predicting Polarities of Tweets by Composing Word Embeddings with Long Short-Term Memory. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 26-31 July, Volume Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), p. 1343–1353.