Natural Language Processing Applications: A New Taxonomy using Textual Entailment

Manar Elshazly[1], Mohammed Haggag[2], Soha Ahmed Ehssan[3]

Computer Science Department

Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt

Abstract—**Textual entailment recognition is one of the recent challenges of the Natural Language Processing (NLP) domain. Deep learning strategies are used in the work of text entailment instead of traditional Machine learning or raw coding to achieve new enhanced results. Textual entailment is also used in the substantial applications of NLP such as summarization, machine translation, sentiment analysis, and information verification. Text entailment is more precise than traditional Natural Language Processing techniques in extracting emotions from text because the sentiment of any text can be clarified by textual entailment. For this purpose, when combining a textual entailment with deep learning, they can hugely showed an improvement in performance accuracy and aid in new applications such as depression detection. This paper lists and describes applications of natural language processing regarding textual entailment. Various applications and approaches are discussed. Moreover, datasets, algorithms, resources, and performance evaluation for each model is included. Also, it compares textual entailment application models according to the method used, the result for each model, and the pros and cons of each model.**

Keywords—*Textual entailment; deep learning; summarization; sentiment analysis; information verification; machine learning; depression detection*

I. INTRODUCTION

Textual entailment[1] is the process of importing a text from another one. Textual entailment [2] is an obtaining the relationship between fragments of text when one fragment in some sense implies the other.

Recognition of text entailment (also called natural language inference) is the task of deciding on the allocation of two pieces of text whether the interpretation of one text can be given in another text. Text entailment is a relationship-related activity and sometimes there are threee relationships can occur between two sentences. Entailment- [1, 3] the meaning of one sentence can be combined with another sentence, incorrect entailment - the meaning of one text contradicts the meaning of another sentence, and non-entailment - the meaning of one text does not include or contradict the meaning of another sentence. Or may be categorized to two class (ENTAILS/NEUTRAL).

Textual entailment research mainly contains methods for developing and evaluating algorithmic identification methods of entailment [2]. Text entailment aims at a deeper understanding of text and thinking, which shares the same type of machine learning comprehension, although the structure of the functions is slightly different [4].

Similar to Semantic Textual similarity, in the task of Textual Entailment, the input also contains two sentences (pre-sentence P and hypothesis H sentence), and programs should determine whether the meaning of sentence H can be extended from the meaning of sentence P. In particular, [5, 6] systems are required to assign each line of the phrase or label of CONFIDENCE (when P comprises H, that is, H cannot be false when P is true), the CONTRADICTION label (when P contradicts H, i.e. H is false whenever P is true), or the INDEPENDENT label (where true of H could be determined on the basis of P).

Textual Entailment task is a problem of classification [5]. TE is the latest and best indicator of text acquisition in NLP applications [7].Recent work on Text Entailment work includes (1) models based on linked elements and machine-learning classifiers, (2) sentence-based models that use similar structures such as Semantic Textual similarity, sentence-based function, and (3) other neural network models that do not use Simple RRNs or CNNs as sentence encoders.

Textual Entailment (TE) is an inconsistent relationship between two expressions in which the meaning of one phrase, called Hypothesis (H), derives from another expression called Text (T). The definition of TE is strong in the sense that if H enters from T but has little or no additional information, then the two are treated as unrelated or non-entailed [8].

Text entailment is used in text encryption methods to measure text connections. TE in NLP is a direct relationship between the sections of the text. Relationships are active and effective when the fact of one text fragment follows from another fragment. In the TE framework, integrated and integrated texts are defined as text (T) and hypothesis (H) [7] .

Textual Entailment (TE) is a typical example of a semantic inference, in which the purpose is to determine where a textual hypothesis (H) can be correctly included in a given text (T). While classifying the hypothesis into genes and making an effort to determine where each is differentiated separately by T is called Partial Textual Entailment (PTE) [7].

Partial Textual entailment (PTE) is a possible solution to this problem that defines the interdependent relationships between T and H pairs. PTE relationships can play an important role in a variety of programs to use Natural Language Processing (NLP) such as text summaries and a question answering system by minimizing unwanted information [8].

Recognition of textual Entailment (RTE) [8] is an important function in Natural Language Processing (NLP) research. The key role of RTE includes a question-and-answer system where the nature of the engagement is confirmed by the given answer regarding the expected answer, a summary of multiple texts where the sentences should be deleted, and a retrieval system where only documented questions are required. However, the old definition of TE has two important implications for the relationship. First, the relationship does not coincide where H should be inserted from T. Otherwise the two pair are non-entailed. PTE [9] is a two-way relationship between two expressions in which one expression is inserted or slightly dependent from another expression. PTE expands Textual Entailment (TE) concept.

The rest of the paper is organized as follows: Section (II) provides a literature review about NLP applications that uses textual entailment, in Section (III) provides a comparison between machine learning and deep learning in textual entailment. The paper concludes in section. The paper concludes in Section (IV) with an overview of future work that it was aimed to be done

II. NATURAL LANGUAGE PROCESSING APPLICATIONS USING ENTAILMENT

Textual entailment is a relationship that obtains between fragments of text when one fragment in some sense implies the other fragment. The automation of textual entailment recognition supports a wide variety of text-based tasks, including information retrieval, information extraction, question answering, text summarization, and machine translation [2].

A. Text Summarization

NLP has several applications, the challenging one is a text Summarization. Summarization is a process of presenting a text's most important and relevant information in precise way. It helps to understand a text very quickly and precisely.

Determining the level of correlation between input sentences will help to reduce the inclusion of meaningless sentences in abbreviated text. The importance of the summary becomes even more important and effective when dealing with events on social media platforms such as Twitter, where information is fast, rich, varied and ever-growing content, no matter where, time and other circumstances.

Any user can post comments or watch, upload videos to any event very quickly. Current methods of measuring sentences only find similarities between words and sentences. These methods only state the correct information for each sentence. Summarizing text helps to understand any major text in a very short time .It is to produce the right summary, there are a few steps to follow, e.g. Lexical Analysis, Semantic Analysis and Syntactic Analysis.

The content of the summary story details depends on the needs of the user. Text Summary Methods can be classified as abstractive and an extractive summarization. The abstractive summarization incorporates key ideas into the document and conveys those ideas in clear natural language. The extractive summarization process consists of selecting key phrases, sentences, paragraphs etc. In the original document and combine those with a short form that maintains a sequential sequence [7].

The main purpose in [10] is to measure sentence similarity, which will help to summarize the text of any language, looking only at English and Bengali. The methods proposed in [10] were extensively tested using several English and Bengali texts, from many online sources. This similarity of sentence is stated. Prior to in-depth analysis the best way to prepare for the operation of both languages used is Unicode.

Model [10] represents a new proposed sentence similarity measuring model for English and Bengali language. Lexical methods have been applied, and a perfect expected result has been found.

Lexical Layer Analysis: The lexical presentation and lexical similarity are major function in lexical layer as shown in the Fig. 1. The model followed by one of the popular deep learning algorithm-Text Rank. Lexical Layer Analysis Model has the following steps as:

- Lexical analysis: brake the sentence into tokens.

- Stop words removal: remove pronoun and article words.

- Lemmatization and Stemming: return all tokens to their entry.

Sentence Similarity: The rating method helps to hear the correlation of words that occur in WordNet times. Send the distance between the two words. Other equations to be used in the same suggested algorithms, in this proposed model use python with NLTK version 3, and WS4J: java API designed for WordNet use [10].

However, Twitter's summary work has garnered a lot of attention from the research world over the past decade as a result of sharing fast and varied responses from millions of users.[9] While, summarization of Twitter event is much harder than traditional texts due to the different type of text that poses many challenges. Like:

- Processing of tweet content : this, standard Natural Language Processing (NLP) tools do not work well in this type of text and create challenges for tweet content processing, design integration and topic by topic due to the high nature of understanding features.

- Sentence Boundary Detection (SBD): A tweet does not consist of a single sentence, instead, it can include multiple sentences. Sometimes, a tweet does not include punctuation even if it includes multiple sentences.
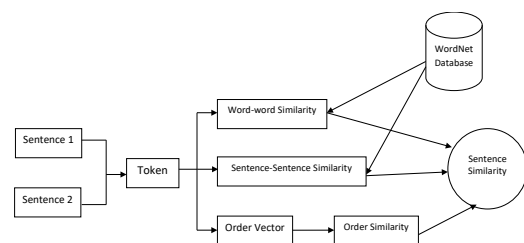

Fig. 1.   Lexical Layer Analysis Model[10].

- Information redundancy problem: Events on Twitter attract a large number of tweets making it difficult to find the most important personal tweets. Every event contains incorrect details of an event in the form of a Re-tweet or titter with the full or partial content of another tweet in high volume. To reduce misunderstandings, identifying partially embedded information between tweets is harder than seeing re-tweets or embedded tweets. For example, two sentences are listed below in two separate tweets. The second sentence contains the information included in the first sentence and additional content (highlighted). To reduce the amount of non-reversible information, these two sentences can be combined into one meaningful sentence as shown below.

Model [9] proposes a summarization approach to extracting an amazing event summary from Twitter by providing a solution that can solve the above challenges. Fig. 2 illustrates this summary. This approach follows a detailed design to produce a summary of the most informative and relevant sentences from all event tweets. After they made this model they found that:

- In many previous researches, tweet is considered a summary of work. But in reality, regular tweets include many sentences. In this research the authors extract important sentences by SBD from tweets instead of treating the tweet as a sentence. In order to exclude the most important events of the event, they put all relevant tweets of the event first and convert the top tweets placed into possible sentences.

- To produce an abstract summary, authors filter out invalid data by checking the Partial Textual Entailment (PTE) relationship between sentences. They see peer-to-peer (PE) sentences and a text border that is slightly inserted between sentences. Couples of PE sentences are combined into one sentence by combining PE text. Finally, an abstract summary creates selected sentences in addition to covering high detail, diversity, consistency and readability.

In [9] the proposed model is two phase summarization approach. First phase is key sentence extractor which produces more relevant and informative sentences rather than tweets from the appropriate tweet event set, second phase is abstractive summarizer which produces a desirable summary over the selected sentences after deleting the unwanted information by pointing to the information included or partially included in the sentences. So, authors have seen the quality of the summary is limited by the accuracy of the TE determination. This [7] seeks to overcome these restrictions by using Partial Textual Entailment (PTE) as well. Hence, [7] suggested a modified approach that demonstrated competitive outcomes in standard dataset.
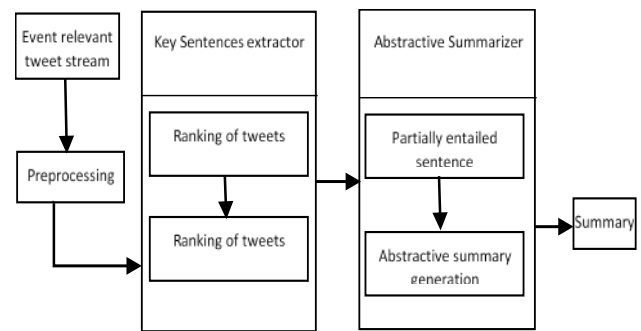

Fig. 2. [9] Summarization Approach.

Therefore model [7] suggests a new and exciting way to summarize one document. It also includes Partial Textual Entailment (PTE) as a way to get the power of expression within a couple of texts. Sentences are divided into units of text for the purpose of TE to compute which is why they are different from complete TE and are called intra-phrase sentences or PTE. Since a graph model is known to capture and present a well-defined structure of information with information shared between interconnected areas within a complex network, so authors will use a graph presentation to illustrate the relationship between different pieces of text. Then authors used the weighted (Minimum Vertex Cover) method of MVC, a graphical algorithm to get a small vertices' set (representing sentences) that make up the cover. And this cover forms their final summary [7].

The text summary function was created as a WMVC problem by them. The input document provided to the system to be compensated and rewritten and labeled as a graph network. They built a targeted, weighted graph model where border instruments could not be considered. Accordingly, the weight w is already expressed in each graph node based on the entailment values obtained. This improved graph form to date is placed under WMVC method [7]. This algorithm works by narrowing down the graphical network so that the cover C is a subset (part) of the smallest sentences weight, meaning the best-placed sentences in the text. The cover that has been found is this result - a summary of the input document. Model [7] Recommended Method of partial Text entailment Used Baseline Weighted Minimum Vertex Cover (PTE-WMVC), shown in Fig. 3 is contained the following key steps:

- Pre-processing step

- PTE score computation

- Salience scores computation

- PTE based graph construction
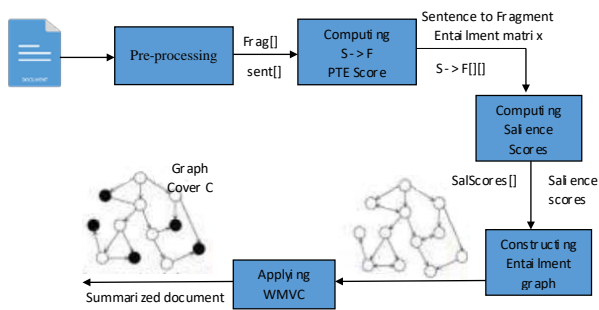
- Applying WMVC algorithm

Fig. 3. PTE-WMVC Structure Diagram [7].

Finally, the authors compared the effectiveness of this approach with other leading methods in the field of text summarization. This PTE-WMVC system provided the highest recall value of 50.99% of ROGUE-1 while n-gram size was considered N = 4. The next best performance is provided by WMVC with a recall value of 38.8% followed is PDT at a rate of 34.6%. A very small amount of recall was identified by the program and ATESC as the basis for text processing in the system [7]. In [7], authors introduced an amazing and advanced way of summarizing a single text. Through extensive experiments in the database that showed in this paper the proposed algorithm uses Partial Textual Entailment at its end it uses existing techniques that look at complete textual entailment to establish relationships between sentences.

The suggested algorithm has been used successfully in small sections of sentences, uses Partial Textual Entailment and produces clear and reliable measurements. Building such a system and dealing with the same method of summarizing multiple documents by taking advantage of the functionality of the proposed algorithm involves future work. Future work of [7] will be more focused on summarizing many documents. The creation of a summary requires multiple textual sources and often involves a combination of collections of various texts.

B. Question Answering

Question Answer (QA) is an ancient research field in computer science that started in the sixties. In the Semantic Web, most of the suspended data is now available in the form of knowledge bases (KBs). Today, there are KBs on media, publishing, geography, life-science and more. The main purpose of the QA program over KBs is to retrieve the information you want from KB or more, using native language questions.

In summary, most QA programs only work in English and more than one KB. Multilingualism is taken for granted while carrying it can often be taken for granted. The fact that QA systems tend to use existing technologies and require several services to be presented to the end user, leads to the idea of developing QA systems in a general way [11].

Model [12] introduced new data and text implant model, which is derived from the treatment of answering multiple queries in the selection as an input problem. SCITAIL is the first input set created only from natural sentences that already exist independently "in the field" rather than sentences written specifically for the input function. In addition to the required

input details, they construct add-ons from scientific questions and related answers, as well as structures from relevant web phrases taken from the larger corpus. These sentences are often a language challenge. This, combined with the high similarity of the baseline and hypothesis in both paired and non-entailed pairs, makes this new entailment work even more difficult.

Model [12] present the largest input data taken directly from the final task and contain text that appears naturally as both the basis and hypothesis. The model new data, SCITAIL, is designed for the final task of answering high school science questions. Each question and their relevant answer is converted into a supporting statement for hypothesis H. they use the retrieval (IR) method to obtain relevant text from the business of large web text documents, and use one of these sentences as Layout P. -H, not all that includes P or "supports" the statement in H. They include annotations for each hypothesis structure head as a support or not, to create SCITAIL input data with 27K examples.

Since both the foundation and the drawings in SCITAIL were written independently of each other and without the inclusion function, the linguistic diversity of the database is not limited to the coverage of handwritten rules or by crowds of clever staff two pieces of text [12].

They show that one can improve SCITAIL accuracy by 5 percent using a new neural model that exploits language structure. And found that current RTE systems, including neural input models, have interoperable functionality in this data, whether they are trained in their data or in SCITAIL. Specifically, their asymmetric Decomposed Graph Entailment Model (DGEM) raises an accuracy by 2.3%. They make the following contributions:

1) A natural input database in which the text and transcripts were written independently of each other and independently of the input function;

2) Initial data taken from the final task of answering multiple questions; and

3) A new model that uses language more recklessly in the hypothesis than the existing strategies in this data.

Recently, text submissions have received significant attention due to their extensive use of retrieval information, commendation programs and answering questions. Information extraction (IE) is an Automatic retrieval of certain information related to a selected topic from the body or bodies of text. Extraction tools enable us to retrieve data from text, information, websites or multiple sources.

Recently, various frameworks for textual entailment recognition have been proposed, ranging from traditional linguistic strategies pertaining to different deep learning methods. However, the deep neural networks that reach the state of the art in textual entailment task look only at the details of the given sentence structure rather than the real world knowledge and information beyond context. In [13], they propose a Knowledge-Context Interactive Textual Entailment Network (KCITEN) that reads graph-level presentations by tying a graph of external information through a graph-focused network. They also propose a Text-graph interactive reading

system for neural learning entailment, which allows for unwanted noise and noise and focuses on informative presentations. Examination in the SciTail database shows that KCI-TEN works better than state-of-the-art methods [13].In addition to the practical application of previous studies, textual entailment recognition models have two limitations as follows:

- Underutilize the large-scale real-world knowledge: Existing RTE models may suffer from language challenges due to the high similarity of the lexical premise and the hypothesis of both blocks included and excluded.

- Perform learning and inference between text and large scale KG without explicit interaction: These methods [13] are useful for easy comprehension of concepts but can lead to a lack of communication between text-based embedding and KG-based embedding during the editing process.

To address these two limitations , in [13] proposing a novel representation framework, the Knowledge-Context Interactive Textual Entailment Network (KCI-TEN), which makes full use of the external KG to determine whether a given text incorporates a given hypothesis. Specifically, the proposed approach begins to convert the concepts and basic sentences into KG sub graph, incorporating them into the context and contexts of the environment through a graph attention network (GAT), providing an additional incentive to explore external sources of textual recognition. Thereafter, the included vectors are used to calculate the basis for comparison-hypothesis corresponding to the text level and graph level respectively. Finally, using a method to read the matching representation in size, using practical information from text and KG, and re-inserted into the classifier to get textual entailment [13]. The model concise contributions are:

- Creating an effective RTE communication framework, which directs the external KG to obtain background information and understanding.

- Investigating the feasibility of graph-based submission by approving GAT to carefully insert KG sub graphs.

The results of the experiments indicate that KCI-TEN successfully incorporates the conditions in terms of KG sub graphs [13].

Recognizing inclusion relationships between sentences is an important and common part of language communication. RTE function has been proposed as the solution to this problem. In [14], they introduce the Arabic Recognizing Textual Entailment Tool called Ar-SLoTE "Arabic Semantic Logical Textual Entailment Tool". Model [14] suggested tool is made up of five modules: pre-processing, linguistic analysis, first-order logic representation, features extraction and entailment decision modules. It produces logical representations of hypothesis / text pairs in order to extract important and informative features, namely, predicting the dynamics of contradictions, exactly the similarities.

Ar-SLoTE is mainly based on Arabic question / answer procedures and the results obtained are very encouraging. Obtaining RTE In many NLP applications is by performing

many tests, including the Question-Answering (QA) programs. Therefore, RTE helps to determine whether the given answer entails the question asked or not. There are various methods suggested in the literature to determine which entailment into this field. However, in the Arabic language, there is a shortage of RTE services provided in QA programs.

As long as, in [14] they are introducing an Arabic tool called Ar-SLotTE «Arabic Semantic Logical Textual Entailment Tool». For defining the entailment between a couple of question and answer they propose this tool. The correlation of the entailment is set as "TRUE" when it means how the question can be estimated from the answer and "FALSE" in another way.

Model [14] is based on the first logical presentations of the structure orderly to extract new semantic information from the H-hypothesis and the T-text. Three factors are excluded from the logical expression of T and H namely, the accumulation of Predicates-Arguments, the Semantic similarity and the similarity of the Name entity. A rating is given for each feature. After that, the feature vector is designed for each H / T pair. These doses are passed on to a tree decision reader to guess the category of each couple: «TRUE» if there is an entailment between them, and "FALSE» in another way. Ar-SLoTE is designed to detect the entailment relationships that exist between Arabic H/T. In particular, it sees the entailment between the true response and its answers. Noting that the real question is a question related to an entity with names.

In this situation, the hypothesis indicates the question asked and the text identifies the given answer. Thus, Ar- SLoTE takes, as an input, questions that start with test nouns (من who,ما what, أين where, متى when) and a list of related candidate responses. At the last, the entailment between the question and each answer was indicated. Research correspondent's response is considered an appropriate answer to a question [14].
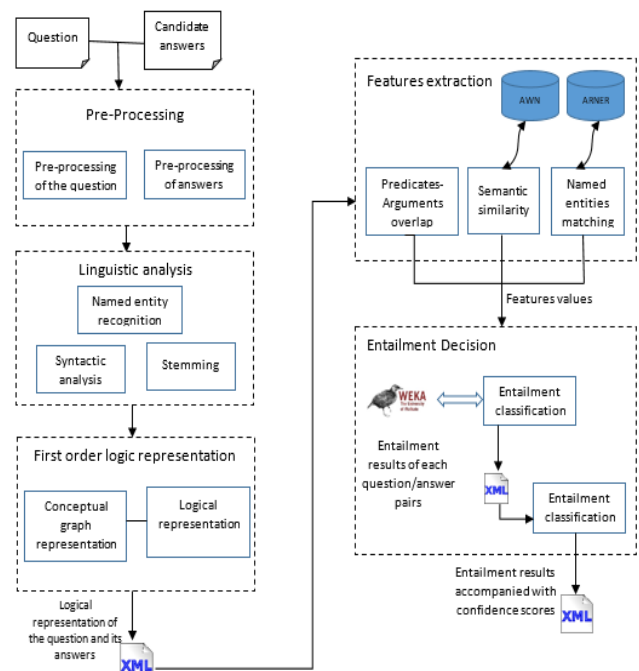

Fig. 4. Ar-SLoTE Architecture [14].

As shown in Fig. 4 Ar-SLoTE architecture is built on five key modules, namely: pre-processing editing, linguistic analysis, first-order logic representation, feature extraction and entailment decision modules.

Model [14] approach focuses on the exclusion of elements from the two logical pathways taken from T and H, specifically, predicting contradictions, semantic similarities and word similarities. The Ar-SLoTE releases a file from type XML containing the accompanied results of entailment using confidence scores. Therefore, the result can be shown and translated by the user, or by machine.

As the [14] tool is designed for question / answer processes, authors have selected 500 operators of a variety of authentic questions and answers from the AQA-WebCorp Corpus. At first, they turned them into logical submissions. Second, the corpus was divided into two data sets: training set (350 pairs) and test set (150 pairs).

Reported to the following TABLE I, the results obtained by the proposed tool in [15], Ar-SLoTE and LR-ALL. Among the state-of-the-art technology that works for Arabic RTE, the tool proposed in [15] is the only function specifically intended for real-time question / answer systems. LR-All [16] Arab RTE system with highest accuracy. This paper uses traditional embedding features and words. In addition, it is based on machine learning and uses ArbTEDS dataset, the only available Arabic RTE data containing 600 pairs of T-H.

They have been able to be more accurate compared to the previous RTE system which was intended for question / answer applications (Ben-sghaier et al., 2018) [15]. Because the results obtained in [14]show that reasonable representation and appropriate relevance can improve the results of Arabic RTE. Authors said that Ar-SLoTE performance can be improved by integrating, in particular, additional cognitive information. One of authors' plans that focuses on testing their RTE tool on a large computer uses a large number of different questions / answers per pair. As a long-term quest, they aim to look at other types of questions [14].

TABLE I.   COMPARISON BETWEEN AR-SLOTE AND OTHER ARABIC RTE SYSTEMS

System	Approach	Dataset	Accuracy
Ben-Sghaier, M. , W. Bakari, and M. Neji [15]	• Machine Learning based . • Use of semantic measure and word sense disambiguation.	200 Question/Answers pairs	70%
Ar-SLoTE	• Machine learning based • Use of logical representation and extraction of features	500 Question/ Answers pairs	73.33%
LR-ALL [16]	• Machine learning based • Use of traditional and word embedding features	ArbTEDS	76.2%

C. Information Verification

As the name itself implies, data verification is a process that verifies the authenticity of a person or company for a particular process that requires verified information. This process helps to verify the current status, including features such as hiring and accommodation, allowing an application that requires specific references to be completed by the preferred person or organization of applications.

Information verification is important because the level of information production is high and growing daily, usually on social media. This also results in social media being included as a mass media. Therefore, data verification in social media becomes even more important [6].

Information verification is a form of journalism investigation in which experts examine claims that have been published by others about their truthfulness. Claims can range from statements made by members of the public to stories reported by other publishers. The ultimate goal of a fact-checking system is to determine whether a claim is true, false or mixed.

In [6], a method of verifying information on Twitter is proposed. The proposed method uses textual entailment methods to improve the authentication methods on Twitter. Separating the effects of entailment methods over state-of-the-art methods can produce tweet verification results. In addition, as the writing style of tweets is incomplete and structured enough for textual entailment, they have used the language model to add tweets with official and appropriate text for textual entailment.

While the methods used to incorporate reinforcement inputs to verify data may have acceptable results, it is not possible to provide relevant sources for all tweets, especially in the past by posting tweets. Therefore, they have used other sources such as User conversational tree (UCT) without using input methods to verify tweet details. UCT analysis is based on pattern release at UCT. Test results show that using entailment methods increases tweet validation [6] as shown in the model architecture in Fig. 5.

Looking at the various challenges in obtaining rumors, they look at the findings of the rumor by the two-source analysis: (1) entailment based classifier, and (2) UCT-based classifier. Then, in each of the two sources, they train two different classifiers separately, and after that, use a weighted ensemble voting classifier, the results of these different classifiers combined to form a new classifier [6].

All in all, contribution to [6] to verify the details on Twitter is:

- Textual entailment is used to confirm rumors on Twitter

- A language model is used to make tweets more acceptable in the style of writing

- Consideration under UCT analysis

- Promote a weighted phase of voting to integrate the results of the inclusion process and UCT analysis
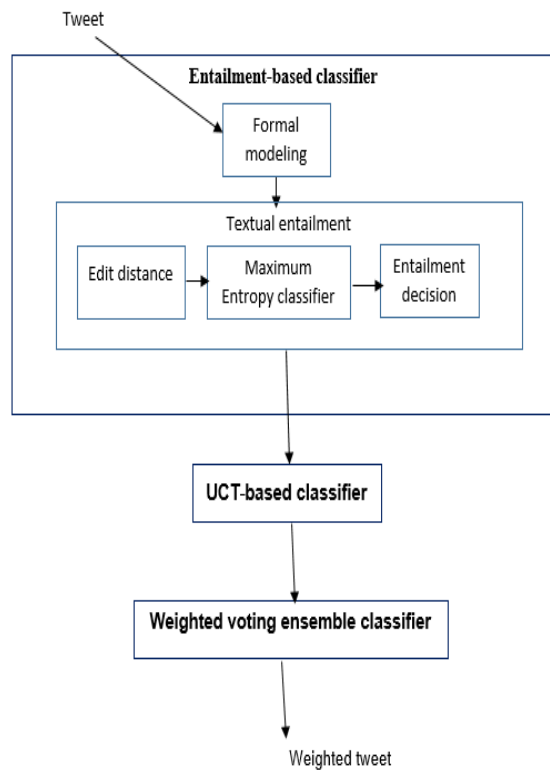
Fig. 5.   Model [6] Architecture.

Information on the web is linked to rumors and unverified information. In addition, social networks as a special and broad part of the web have a lot of potential for spreading and creating inaccurate or unauthorized information.

Because of the importance of this issue, as well as improving the effectiveness of data verification, in [17] verification of data on social media is being investigated. Several features and conditions appear to be applicable to the detection of rumors.

Among the functional and structural features, authors consider two important sources of data validation in social media that include user responses and news organizations. User feedbacks as the primary source can be a user conversional tree. Some patterns can be drawn from this tree. Media agencies as a second source are also used for data verification by tools of text entailment. Finally, these two types of factors are combined to separate information into one of three categories true, false, or unconfirmed. This method was tested by checking public data sets. Test results show that the proposed hybrid information verification method can surpass highly recommended methods in data verification [17].

Today, everyone has some doubts about the information broadcast on the web or social media. This is because the amount of information that is visible on the web is wrong or unreliable. Moreover, people are exposed to all kinds of fake and uncertain information. Doubts about the facts of the information available on social media create public concern, especially in difficult situations or disasters. A large number of online false stories have the potential to cause great social problems. Stock markets has been affected by fake information

on the web and social media, reduced responses during disasters and terrorist attacks [17]. Typically, bots collaborate to create and disseminate fake or wrong information on a large scale using bots accounting groups that work on a large scale for two purposes:

- by distributing the same content, e.g., by retweeting, to multiple viewers, and

- Increase social status by following each other's accounts and following trust information [17].

Those who try to stop the spread of wrong information are working to create programs that can reduce the way bots and other programs spread lies and slander. Most of the previous methods depend on the selection of the feature from the main text. In addition, the methods studied in the structure of graphs in social media [17] are popular with the rich history of graph reading. Recently, more comprehensive learning approaches have also emerged in this field. For example, recurrent neural networks (RNNs) are used to detect rumors from microblogs.

However, among current methods, the effect of resources on data verification cannot be ruled out. Therefore, in [17] to strengthen the validity of the data verification, the results of the sources are studied. The proposed method of this data in [6] uses the main sources of information, mainly feedbacks, and media sources. Initially, the effect of user feedback and media on the data verification function are studied separately.

Subsequently, these methods' results are gathered to determine the validity of the information. In the user feedbacks, the composition of the marked areas in the user conversational tree (UCT) by four types of tags, rejected, comment, supporting, and query, is considered. UCT is being studied discussed in the following three ways:[17]

- Pattern extraction: For UCT pattern extraction, the unique UCT patterns are calculated by looking at the pattern level in the tree. In these patterns, a tag for each node is selected, too.

- Statistical sequential models: By studying UCT sequencing models, two different types of sequence models are calculated

- Hidden Markov model (HMM)

- Conditional random fields (CRF).

- Edit distance: Set UCT learning methods used as standard K-nearest approaches to UCT classification with data verification function.

The second source of data verification is the media sources, which is useful for textual entailment methods. At this opinion, the text that feels authentic (called the main text) is considered to be either media or not. If not entailed, the main text is considered authentic. When the media compares with the main text, the main text is considered false. Otherwise, the main text is considered unverified. Suggested data verification methods are evaluated using public records by comparing the state-of-the-art-methods.

The results of these comparisons show that the proposed method works better than the existing alternatives presented in [17] database used.

The description of the method starts with the method's full description in the first subsection. After that, the two main parts of the method are explained, that were namely:

- Studying of user feedbacks and various components

- News sources studying

An overview of the proposed [17] approach to data verification activities is shown in Fig. 6. As shown in the figure, the input for data validation is a message from a social network.

After reviewing the input text, you are ready for verification. Using two different steps, verification work will be done: it processes feedbacks of users and new sources.

1) User feedback study: User response research is typically based on the UCT structure. UCT is a tree created when a message is communicated on social media or is often responded to. The root of this tree is the main message its authenticity is being questioned. After all, every response to this message is a child root. Previously, every response of each node in a tree is the next level of UCT.[17]

Thereafter, the UCT tagged is used to predict the final label of the input text using the following three methods in the next three paragraphs. UCT is studied in three ways including:

2) Pattern extraction: UCT derived pattern are considered features. These patterns are below each starting point with a maximum of three. Calculation of each pattern's occurrence is by looking at the weight, which is calculated by the level of the root scale relative to the height of each UCT.

3) Statistical sequential models: HMM and CRF are used in [17]proposed methods. The path label with the most voted priority is false, true, and unconfirmed is considered the last label.

4) Edit distance: Set the distance of each path at UCT calculated the most effective dynamic time strategy. this method uses a powerful planning method [17].

5) Study news sources: recognizing textual entailment

In this case, a source of information should be provided. When the main text, in which the message is confirmed as true message, is entailed from the new source, the source text is the first text. If the new source differs from the main text, then the main text is fake or wrong. Otherwise, the main text is not verified. The following methods of textual entailment are used in [17]:

- Edit distance comp (Fixed Weight Lemma/RES Word Net).

- MaxEnt Tree Skeleton RES (Verb Ocean Tree Pattern/ Word Net Tree Pattern/Word Net Verb Ocean Tree Pattern).

- P1EDA RES: Paraphrase Table.

To compare the [17] method with the state-of-the-art methods in which authors simulate tests similar to those performed in the Semeval-2017 work 8. The test methods are similar to this work, too. Test methods are Score, Confidence, Root mean Square Error (RMSE), and Final Score. The Score is ranked as the most known accuracy measure. Confidence RMSE is the RMSE Classified Confidence.

Existing new sources are critical to the success of this approach. And on some social networks, user feedback is not recorded and in this case this method is bad choice for data verification. In addition, the [17] method only applies to text messages. Examination of the this method of data verification in public dataset showed that the [17] method surpassed those who are technically competent in data verification.

Model [18] introduces ColumbiaNLP's submission of the FEVER Workshop Shared Task. This model is an end-to-end pipeline that delivers factual evidence using Wikipedia and provides a decision on the veracity of the claim based on evidence-based evidence. The FEVER shared function aims to test the capabilities of the data verification system using evidence from Wikipedia. Given a claim that includes one or more items (mapping to Wikipedia pages),the system in [18] must take written evidence (sentence sets on Wikipedia pages) supporting or disputing the claim and using this evidence, they must write a claim as Supported, Refutes or Not_Enough_Info. The shared work dataset is presented by [19] and contains 185,445 claims. *TABLE II* shows three scenarios from a set of claimant data, evidence and decision. The first system that used by [19] uses 3 main components:

- Document retrieval: using the documentation recovery component from the DrQA system that retrieves the adjacent text query using the cosine similarity between attached unigram and the bigram Term Frequency- Inverse Document Frequency (TF-IDF) vectors.

TABLE II.   THREE SCENARIOS FOR CLAIM AND EVIDENCE AND DECISION OF [19]

Claim : Claim : Fox 2000 Pictures released the film Soul Food. [**wiki/Soul_Food_(film)**]   **Evidence**: Soul Food is a 1997 American comedy-drama film produced by Kenneth "Babyface" Edmonds , Tracey Edmonds and Robert Teitel and released by Fox 2000 Pictures .   **Verdict**: SUPPORTS
Claim : Murda Beatz's real name is Marshall Mathers. [**wiki/Murda_Beatz**]   **Evidence**: Shane Lee Lindstrom (born February 11, 1994), known professionally as Murda Beatz, is a Canadian hip hop record producer and songwriter from Fort Erie, Ontario.   **Verdict**: REFUTES
Claim : L.A. Reid has served as the CEO of Arista Records for four years. [**wiki/L.A._Reid**]   **Evidence**: He has served as the chairman and CEO of Epic Records, a division of Sony Music Entertainment, the president and CEO of Arista Records, and the chairman and CEO of the Island Def Jam Music Group.   **Verdict**: NOT ENOUGH INFO

- Sentence Selection: [19] simple sentence selection method sets out the similarities between TF-IDF and claim. They set the most similar sentences first and then using validation accuracy on the development set, they accurate a cut-off. Then they test both DrQA and the easy use of the unigram TF-IDF to measure selection sentences. They also examined the effect of sentence selection on the RTE module by predicting the entailment given to the original texts without sentence selection.

- Textual Entailment: It is a multi-layer perceptron (MLP) with a single hidden layer that uses frequencies of term and TF-IDF cosine similarities between claims and evidence as features. To analyze the state-of-the-art in RTE, they have used the decomposable attention model (DA) between the claim and the role of evidence.

Overall, [18] end-to-end model shows an improvement of 19.56 on FEVER results compared to the previous [19] (50.83 vs. 31.27) in the development set. system [18] of utilizes changes in all modules that lead to significant improvements in both development and testing sets:

- Document Retrieval: this step is an important step in establishing an end-to-end authentication and verification system. Missing the correct text can lead to untranslated evidence, while incorrect text can add to the noise of subsequent sentence selection task and textual entailment task [18] . In [18] authors suggest a multi-step approach to retrieving documents appropriate to claims:

  o Google Custom Search API: inspired by previous research [20], the Custom Search API of Google was used first to retrieve documents having information about the claim. The token Wikipedia was added to the claim and produce a query and gather the top 2 results.

  o Named Business Recognition: Identifying an entity name using a previous activity, and after finding the entities that be named in the claim, authors use the Wikipedia python API 3 to gather a high-rated wikipedia document returned by the API for each named entity.

  o Dependency Parse: to increase the chance of finding the right entities in the claim, they get the first sentence of lowercase letters, Reason to emphasize the lower case verb phrase to avoid non-claim entities such as "Finding Dory was directed by X", where the appropriate entity is "Finding Dory" [18].Then they decide to create better accounting models to handle entity ambiguity or entity linkages in the future.

  o Combined: Authors use the documentation union returned by these three methods as the final set of relevant documents to be used by the Sentence Selection. They note that their (combined) method in [18] reached a high accuracy of 94.4% compared to the original

method [19] of 55.3%. Because authors do not have best evidence of blind set test they cannot report claim coverage using their pipe.

- Sentence Selection: authors used a component to retrieve translated texts of DRQA [20] to sentences selection using the bigram TF-IDF for binning as suggested by [19]. They extract the top five sentences in the appropriate text k-most the related documents by using TF-IDF vector matching. To solve the previous problem that arose in selecting sentences of [19] authors continued to filter out the top 3 evidence from chosen 5 evidences using distributed semantic representations. Model [21] show how the deep representation of words modulates the complex aspects of word use (e.g., syntax and semantics), and its use in a variety of language contexts. Therefore, converting claim and evidence to vectors using ELMo embedding. then [18] calculated the cosine similarity between the claim and the vector of evidence and produced the top 3 sentences based on their score.

- Textual Entailment: authors did not show evidence, but trained model for each pair of evidence. For textual entailment recognition they used the previous model in their work on the supervised learning of the representation of universal sentences. They use bidirectional LSTMs for max-pooling to encode a claim and evidence. In word matters in pre-trained word embedding it is a major stumbling block to sentence representation. To resolve this they use fast text embedding based on subtitle information. Also, this embedded training in the Wikipedia corpus makes them the right choice for this method. Their final predictions are relied on the next algorithm in Fig. 7.

Because the selected evidence was naturally noisy and this pipeline [18] did not verify the evidence together they chose this rule rather than more prediction to reduce the rule of prediction of the NOT_ENOUGH_INFO class.

```
Algorithm 1 :
if count(Support)=1 and counr(Not_Enough_Info)=2 then
print(Support)
else if count(Refutes)=1 and count(Not_Enough_Info)=2 then
print(Refutes)
else
print(arg max((count(Support), count(Refutes),count(Not Enough Info))
```

Fig. 6.   Algorithm Used To Make A Decision In Module.

TABLE III.   FEVER SCORES ON SHARED TASK DEV AND TEST SET

Data	Pipeline	FEVER
DEV	[19] module	31.27
	This module	50.83
TEST	[19] module	27.45
	This module	49.06

Module [18] also tried to train a classifier that takes confidence scores from all of the claim/evidence and their status in the text and trained an improved classifier but the accuracy was not improved. Experimentally the rule has given them good results in the development set and thus they used it to get the final label. *TABLE III* shows the total FEVER points obtained by [18] pipeline in development and test sets. In that time ranking [18] system ranked 6th.

The new models cannot rely on models that rely entirely relied on semantics. Although the two sentences are similar, the similarity of the cosine between them is worse because the evidence contains many additional details that may be inconsistent with the claim and difficult for the model to understand. They also found cases where the predicted evidence was accurate but not consistent with best evidence. But this system is being penalized on not being able to match the best evidence. Module [18] has found a few annotations that are incorrect which is why FEVER scores are lower than expected. Sometimes, the lines between SUPPORT and NOT ENOUGH INFO are not cleared enough. This models need a better understanding of semantics in order to be able to distinguish between them in all cases.

D. Sentiment Analysis

Sentiment analysis (emotional AI or mining opinion) refers to the use of natural language correction, text analysis, related languages, and biometric to directly identify, extract, measure, and read the corresponding regions and information.

Sentiment Analysis tools now are slow, and it is the most difficult statistics that make the work well done. Sentiment analysis can be subdivided on the basis of a separate textual text. It can be categorized at three levels discussed briefly below:

- Document Level Analysis: analyzes the contents of the document as a whole to determine its availability.

- Speech Level Analysis: Pays attention to the various sentences of a document by continuing to break it into small word texts to analyze its structure.

- Word Level Analysis: Determines the minimum number of different words in a sentence in relation to an item or event.

Previous work defines hypothetical analysis as research to determine emotional AI. Sentiment analysis work involves finding out the feelings, status, person's opinion, object, product or event that has worked in areas such as academics, business and industry. Sentiment analysis work can be categorized on the basis of different classes - can be analyzed as positive, negative or neutral based on class.

Appropriately, authors cannot accurately measure guidance using TE for a number of reasons such as complexity of sentences and limitations of available language resources. Therefore, (PTE) is using to measure the degree of inclusion whether the text fully integrates the concept or not. The authors plan to use Partial Textual Entailment (PTE) to improve the work of Sentiment Analysis in [22]. It is expected to reduce the amount of work done in analyzing the text view thus providing a better solution for SA work. The method proposed in [23] is to carry out the task of analyzing the sentiment by exploiting the concept of partial text entailment. They suggest using the partial textual entailment to measure semantic similarities between tweets in order to combine the same tweets. The method is expected to reduce the burden of sentiment analysis and make processing faster. Then, authors propose to change the method of partial text entailment that can be adapted to most NLP applications. The purpose of [22] is improvement the task of sentiment analysis using partial textual entailment.

The authors propose a two-part Sentiment analysis approach:

a) A method to improve the PTE recognition process and

b) To improve the Sentiment analysis method using the partial textual entailment.

E. Designing a Method to Improve the PTE Recognition Process

The proposed idea is to expand the add-on model to increase due to the large number of feature builds and produce relevant results that done by deleting helpful words. Elements will be constructed only using the sentence title as the key word in the construction of the elements. This approach is expected to predict the best of text input as the text revolves around the topic of the sentence[22].

1) Choose text and Hypothesis

2) Re-analysis of the text and Hypothesis by removing helpful words

3) Subject identification: Subjects from sentences are selected from the remaining words written separately.

4) Structure construction: by matching the remaining words with the subjects

5) Composition of partial text entailment will be performed

F. Creating a Way to Improve Emotional Analysis using Partial Text Entailment: Authors Discuss the Entire SA Framework using PTE Below

1) Recognizing new models or methods for Partial Textual Entailment: this step will do the job of collecting similar tweets to reduce the burden of sentiment Analyzer which has led to the rapid implementation of updates. Integration recognition will be done using the BIUTEE tool.

2) The result obtained in Step One will be forwarded to a mood analyst to get the feelings of the speech.

The total number of sentences that entail the hypothesis will then be provided as an estimator of how strong the sentiment is[22].

Model [23] aims to increase the accuracy of the entailment of Arabic texts using resolving ignoring of the text-hypothesis pair and determining the splendor of the text-hypothesis pair whether it is Positive, Negative or Neutral. It is noteworthy that the absence of a negation detection factor provides negative results when detecting the correlation of the factor since contradictory returns true.

Contradictory words are considered termination words and are removed from the text-hypothesis pair which can lead to improper entailment decision. Another case that has not been resolved before, it is impossible for the correct text to include the wrong text and also the same for the opposite order. In [23], in order to distinguish the text of the hypothesis pair contradiction, a sentiment analysis tool is used.

Authors show that processing the essence of a text-hypothesis pair increases the accuracy of the entailment. To test [23] method of the ArbTEDS archive (ArbTEDS) dataset which comprises 618 text-hypotheses pairs and shows that the accuracy of Arabic entailment increases by resolving the inconsistencies of the entailment and analyzing the unity of the hypothesis pair. The main problem with the original Arabic texting systems is that they do not see the negation, where negation brings back the truth and does not take the critical side where negative or positive thinking does not convey the opposite feeling. To their knowledge, [23] represents the first attempt to recognize the influence of resolving negation and emotional polarity in the Recognition of Textual Entailment in the Arabic language. They conducted research on the Arab Text Entailment Dataset (ArbTEDS) [24]; shows that the more accurate the discovery obtained by resolving the negation and the more consistent finding of this text-hypothesis.

In this work, authors face the challenge of creating an Arabic text-entry system that produces relevant results. The proposed project in [23] ,that it is shown in Fig. 8,is called Sentiment Analysis and Negation Resolving for Arabic Text Entailment (SANATE) which enhances ATE method (ATE Method is described in [23]).
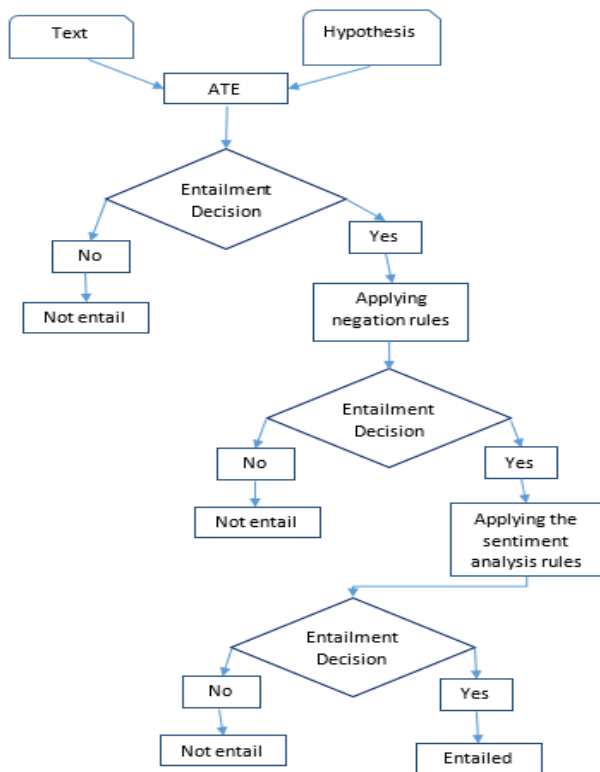

Fig. 7.   General Diagram of Sanate System [23].

To improve ATE, they set some rules for negation. If the decision to include ATE is "not entail" then no further consideration is included. If the decision to include ATE says it will be "entail" then a set of rules will be considered. The rules used to do some text and hypothesis entailment include [23]:

1) If negative particles appear before the same action (normal action) in the T or in the H, the judgment will be: not entails

2) If the negative adverbs appear before the normal action in the T and in the H (both), the judgment is: entails

3) If T and H have a different verb and are negative particles appearing before the verb in the T or in the H or in both, the judgment is: NOT entails

4) If the text-hypothesis pair has more than actions (verbs), then:

a) If one of the normal actions is ignored by one of the negative particles in T or H but is not ignored in both (T and H), the judgment is not entail.

b) If the normal action is ignored by one of the negative adverbs in T and H (both), judgment will be entail.

5) If text and hypothesis contain Opinion words (R):

a) If the result R is not the same as T and H which means different variations, the judgment is not entail.

b) If output R is the same for T and H, judgment will be entail.

To analyze how they use it they have used the 618-text-hypothesis archive (ArbTEDS) database. Each pair of ArbTEDS text is integrated into the ATE system and the SANATE system. Calculation of the accuracy for each system is made by the accuracy of the equation: the measure of the determination of the entailment target to the total value of the entailment problem. ATE accuracy is 0.617 and SANATE accuracy is 0.693. From the impact of the ATE and SANATE system, it is shown that resolving disregard and dividing a text by its contradiction by analyzing emotions, improves the effectiveness of co-identification and non-entailment relationships[23].

Authors conducted research on ArbTEDS; which shows that it is a more accurate result to find entailment to negation and to analyze the unity of this text-hypothesis pair. Without resolving the entailment decision to engage in negation authors may argue, because negation gives the opposite of reality. Some texts may have entailed but the presence of negative particles alters the decision of entailment relationships [23].

Another result from this experiment, finding text and hypothesis pair polarity has a significant impact on finding entailment related to non-entailment. It is impossible for a positive idea to combine negative ideas with the opposite[23].

Fundamental topic in psychology is understanding what makes people in a good mood .Previous work has focused on the development of individual reporting estimation tools and relies on experts to analyze estimated reports from time to time. One of the objectives of the analysis is to understand what is needed to promote individual behavioral change in order to improve the well-being of all [25].

In [25], they set an integral approach; with the view that the user incorporates his or her enjoyable moments as short texts, the system can analyze these texts and provide stable user suggestions that may lead to general improvements in his or her well-being. Authors set one required part of such a program, the Happiness Entailment Recognition (HER) module, which takes as an insert a brief description of the event, a candidate proposal, and outputs the decision of whether the suggestion may be appropriate for this user depending on the event specified.

This part is used as a neural network model with two encoders, one for user input and one for sustainable suggestions, with additional layers for capturing psychologically important features during fun and suggestion. The achievement of [25] the AU-ROC of 0.831 and exceeds their base and current AllenNLP Textual Entailment model by more than 48% development, ensuring the distinction and complexity of HER work.The HER module is described in Fig. 9. This module (1) detects candidate suggestion, and (2) identifies which suggestions are reasonable for specific users. As an input in this module, users create short-term journal entries for "good times," proven interventions to improve subjective well-being.

The first task defines proposals as sustainable in the sense that they promote short-term prosperity without harm in the long run, even if the same suggestion is repeated. For example, a user may report that receiving a puppy or buying a new car has made them happy. However, they do not want to recommend doing them again because repetition is impossible. On the contrary, activities such as swimming or walking in the park have the potential to improve the well-being of the user and are repetitive, These activities were viewed as probable suggestions [25]. Such suggestions become part of this DB Proposal. Appropriately, field experts (e.g., psychologists) would be able to create those consistent suggestions, but [25] also develop a classifier that reflects exciting moments with sustainable activities, and ensure that the classifier accomplished AU-ROC of 0.900.

Second task present the problem of the happiness entailment recognition (HER), which is motivated by the problem of recognizing textual entailment (RTE) in the NLP: given a set of fun times reported by the user, and a set of possible activities, which activities can make the user happy? Model [25] is developing an improved neural network (NN) model with information about concepts, agency, and personalities to predict with the AU-ROC of 0.831 whether the suggestion matches the moment of happiness.

"Concepts" represents topics that are common to HappyDB as an important vector. "Agency" and "personality" are binary indications that the author was in control of their happy time, and that there were other people involved, respectively. They say that this classification includes psychological ideas that are important to that task [25]. The results of the experiment [25] indicate that in this task, the proposed approach exceeds the recognition model for the textual entailment, as well as their foundation. The results support that the HER function is very different from normal text entailment task, and that adding a mental view of the model improves performance.
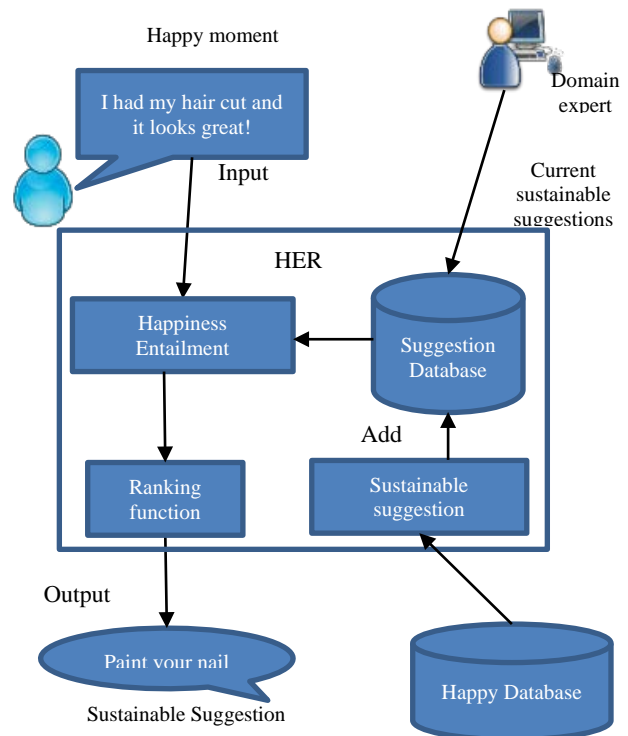

Fig. 8. Her Module and its Inputs and Outputs [25].

III. MACHINE LEARNING VS. DEEP LEARNING IN TEXTUAL ENTAILMENT

A. Machine Learning

Machine learning is used at different stages in the RTE process (i.e. pre-processing, alignment and decision steps), which is usually combined with previous methods. A general model presents this task as a classification problem that is supervised by two classes (entails, not entail) or three classes (entails, not entails, contradiction). Each pair (text, hypothesis) to check is represented by a feature vector, which includes the scores of several operations applied to the pair at different levels[26].Some of described papers used Machine learning to enhance their models . These techniques are Support Vector Machine (SVM), Glove Embedding, Extreme Learning Machine (ELM), decision tree classifier, Naïve Bayes, and Root Mean Square Error (RMSE). All of these techniques are really enhancing the papers result but still in small datasets, and these results compared to deep learning, deep learning enhanced the model in results and with bigger datasets was detected as shown in next table.

B. Deep Learning

Deep learning strategies are becoming increasingly popular in the work of entailment text, overcoming the complexity of different traditional models with complex alignment and ideas. Deep learning in neural networks has become a popular method of machine learning due to the recent success of computer vision, speech recognition, and other areas. This latest achievement is a direct result of being able to train in big data sets, from labeled images for object recognition to compatible machine translation texts. While such information

has been widely available from public sources to date, confidential information collected from individuals will not only provide incentives for existing applications, but also provide new impetus for deep learning. Deep learning is a key technology behind the most recent applications. In deep learning the achievement tasks that directly detect text, image or sound can be learnt by computer models. High accuracy was achieved in deep learning models, sometimes exceeding human-level performance. Textual Entailment approaches will help in many of natural language processing applications, including question answering, summarization, text generation, machine translation, and information extraction. In previous described papers that used several deep learning techniques as Text rank, long short-term Memory (LSTM), Multilayer perceptron (MLP), bi-directional Gated Recurrent Unit (bi-GRU), and Graph Attention Network (GAT).these techniques results with shown next.

TABLE IV. TEXTUAL ENTAILMENT PAPERS

Paper	ML vs DL	Method Used	Comments
Sentence similarity estimation for text summarization using deep learning[10]	DL	Text Rank	Use it in sentence similarity
A New Approach for Twitter Event Summarization Based on Sentence Identification and Partial Textual Entailment[9]	ML	Rank SVM	Use Word2vec to measure word-to-vector similarity of a word give high score compared to another techniques during Rouge metrics. word2vec model is not a deep learning model. It just a library in deep learning
Text Summarization using Partial Textual Entailment based Graphs[7]	No		Used only for single document with small fragments of sentences
A textual entailment dataset from science question answering[12]. DGEM	DL	LSTM MLP	to capture the semantics compute the representation for each node This model raises the accuracy to 77.3%.
Knowledge-aware Textual Entailment with Graph Attention Network [13]	Both	GloVE embedding Bi-GRU GAT	Bi-GRU to encode the contextual patterns Using GAT to encode node contexts in various and noisy KG subgraphs instead of using fixed pre-trained embedding for nodes. Its technique raises the higher accuracy of previous method by +0.8
Ar-SLoTE: A Recognizing Textual Entailment Tool for Arabic Question/Answering Systems [14]	ML	Weka DT	Uses a small set than ArbTEDS, and their model get less accuracy than LR-ALL, but compared to another module its accuracy exceeds another module with 3.33%
Information verification improvement by textual entailment methods [6]	ML	ELM	They want to study more special pattern of UCT ELM is used in deep learning networks but researchers argue if it is considered as a deep learning technique by itself or not even it has only one hidden layer as described before
Information verification in social networks based on user feedback and news agencies [17]	Both	ELM Naïve Bayes SVM RMSE	MLP is used in deep learning techniques but it is not considerd as DNN because it only forward direction Work better than another state-of-the-art-merthod that use the same dataset. The results of hybrid method show that it is more efficent than the results of separate use of user's response approach and entailment approach individually.
Robust Document Retrieval and Individual Evidence Modeling for Fact Extraction and Verification [18]	Yes	Bi-LSTM	Blind set, score = 49.06 > baseline's score= 27.45 dev set , score =50.83 > baseline's score = 31.27
Sentiment Analysis using Partial Textual Entailment [22]	No		This method is not implemented yet , using extended The BIUTEE for RTE
A study of the effect of resolving negation and sentiment analysis in recognizing text entailment for Arabic [23]	No		It has a major cons where it can detects polarity of a text and hypothesis pair significantly affects the detection of entailment and non-entailment
Happiness Entailment: Automating Suggestions for Well-Being [25]	Yes	Bi-LSTM	Bi-LSTM = 0.900 is more accurate by 0.133 with the best accuracy (Logistic Regression)

C. *Difference between ML and DL[27]*

- Data dependencies : Machine learning can trace on less data where deep learning requires large data

- Hardware dependencies : Machine learning can be traced on normal CPU but deep learning requires high performance machines with GPUs to be trained probably

- Feature processing: In ML, most app features must be determined by an expert and entered as data type. Values, shapes, textures, locations, and orientations are features of Machine learning. The performance of many ML algorithms depends on the accuracy of the extracted features. Attempts to obtain high-quality features directly from the data are significant differences between DL and traditional machine learning techniques. Therefore, DL minimizes attempts to design a feature to extract for each problem.so DL has higher accuracy than ML

- Problem solving method : machine learning use many sub-problems and solve them where deep learning is direct end-to-end problem solving

- Interpretability: DL can be tuned in various different ways, while ML has limited tuning capabilities.

- Execution Time: Deep learning in general takes a long time to be trained where some machine learning algorithms have short training and test times.

From the previous comparison, we concluded that deep learning performs better than machine learning. Unlike machine learning, deep learning has proven its ability to handle massive amounts of data. Deep learning can also solve problems that machine learning cannot provide, giving higher accuracy and reducing time. *TABLE IV* shows the textual entailment papers comparison that clarifying method used and whether machine learning or deep learning is applied, including comments on the performance of each method used.

IV. Conclusion and Future Work

Textual entailment is a remarkable field of natural language processing that is used in a variety of applications. Textual entailment applications include text summarization, information validation, question answering, sentiment analysis, etc. This paper spots on famous applications in Textual entailment. By studying previous methods and applications, this survey concludes that deep learning techniques can be used in entailment applications (such as rumor detection and depression detection) to facilitate and enhance the entailment process. It is also concluded that the textual entailment helps to extract data from social media applications and can deal with the huge amount of data more accurately with the aid of deep learning thus, the results are higher in accuracy. When dealing with massive data, more noise and detailed data are found. One text can contain several meanings, and different sentences can show the same concepts. This variation in semantic expression can be considered double trouble of language ambiguity. Textual entailment is the same but minimizes weakness the relationship to be one way. In addition, textual entailment can

achieve text understanding in question answering applications. Despite using question answering applications in e-learning, it can be used in creating a clinical dataset by extracting information from regular documents. Also, it can be used to offhand extract patients' data and export it to health insurance applications.

In future, the GRU technology can be used along with textual entailment. Also, a pipeline containing more algorithms can be used. The aim is to automatically fine-tune the algorithms' parameters to achieve a higher percentage of accuracy and confidence, and discover which of those algorithms can fit the used dataset. Another future direction is to combine NLP and textual entailment with deep learning to outperform results in depression detection in social media. In conclusion, considering the promising results when using GRU in knowledge-aware applications, it is hoped to get higher accuracy when using textual entailment in depression detection.

References

[1] S.H. Jayasinghe, K. Sirts, Deep learning textual entailment system for sinhala language, Publisher, City, 2019.

[2] D.Z. Korman, E. Mack, J. Jett, A.H. Renear, Defining textual entailment, Publisher, City, 2018.

[3] P. Kapanipathi, V. Thost, S.S. Patel, S. Whitehead, I. Abdelaziz, A. Balakrishnan, M. Chang, K.P. Fadnis, R.C. Gunasekara, B. Makni, Infusing Knowledge into the Textual Entailment Task Using Graph Convolutional Networks, in: AAAI, pp. 8074-8081, 2020.

[4] Z. Zhang, Y. Wu, Z. Li, H. Zhao, Explicit contextual semantics for text comprehension, in: Proceedings of the 33rd Pacific Asia Conference on Language, Information and Computation (PACLIC 33), 2019.

[5] L. Zhang, Neural Network Models for Text Understanding, in, 2019.

[6] A. Yavary, H. Sajedi, M.S. Abadeh, Information verification improvement by textual entailment methods, Publisher, City, 2019.

[7] M. Kaur, D. Srivastava, Text Summarization using Partial Textual Entailment based Graphs, in: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) , pp. 366-374, IEEE, 2019.

[8] D. Rudrapal, A. Das, B. Bhattacharya, Recognition of partial textual entailment for Indian social media text, Publisher, City, 2019.

[9] D. Rudrapal, A. Das, B.J.C.y.S. Bhattacharya, A New Approach for Twitter Event Summarization Based on Sentence Identification and Partial Textual Entailment, Publisher, City, 2019.

[10] S. Abujar, M. Hasan, S.A. Hossain, Sentence similarity estimation for text summarization using deep learning, in: Proceedings of the 2nd International Conference on Data Engineering and Communication Technology, Springer , pp. 155-164, 2019.

[11] D. Diefenbach, A. Both, K. Singh, P. Maret, Towards a question answering system over the semantic web, Publisher, City, 2020.

[12] T. Khot, A. Sabharwal, P. Clark, Scitail: A textual entailment dataset from science question answering, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[13] D. Chen, Y. Li, M. Yang, H.-T. Zheng, Y. Shen, Knowledge-aware Textual Entailment with Graph Attention Network, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2145-2148, 2019.

[14] M. Ben-Sghaier, W. Bakari, M. Neji, Ar-SLoTE: A Recognizing Textual Entailment Tool for Arabic Question/Answering Systems, in: 2019 7th International conference on ICT & Accessibility (ICTA), IEEE, pp. 1-6, 2019.

[15] M. Ben-Sghaier, W. Bakari, M. Neji, Recognizing Textual Entailment for Arabic using semantic similarity and Word Sense Disambiguation, in: LPKM, 2018.

[16] N. Almarwani, M. Diab, Arabic textual entailment with word embeddings, in: Proceedings of the third arabic natural language processing workshop, pp. 185-190, 2017.

[17] A. Yavary, H. Sajedi, M.S. Abadeh, Information verification in social networks based on user feedback and news agencies, Publisher, City, 2020.

[18] T. Chakrabarty, T. Alhindi, S. Muresan, Robust Document Retrieval and Individual Evidence Modeling for Fact Extraction and Verification, in: Proceedings of the First Workshop on Fact Extraction and VERification (FEVER) , pp. 127-131, 2018.

[19] J. Thorne, A. Vlachos, O. Cocarascu, C. Christodoulopoulos, A. Mittal, The fact extraction and verification (fever) shared task, Publisher, City, 2018.

[20] D. Chen, A. Fisch, J. Weston, A. Bordes, Reading wikipedia to answer open-domain questions, Publisher, City, 2017.

[21] M.E. Peters, W. Ammar, C. Bhagavatula, R. Power, Semi-supervised sequence tagging with bidirectional language models, Publisher, City, 2017.

[22] S. Gupta, S. Lakra, M. Kaur, Sentiment Analysis using Partial Textual Entailment, in: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, pp. 51-55, 2019.

[23] F.T.J.a.p.a. AL-Khawaldeh, A study of the effect of resolving negation and sentiment analysis in recognizing text entailment for Arabic, Publisher, City, 2019.

[24] D. Majumdar, P. Bhattacharyya, Lexical based text entailment system for main task of RTE6, Publisher, City, 2010.

[25] S. Evensen, Y. Suhara, A. Halevy, V. Li, W.-C. Tan, S. Mumick, Happiness Entailment: Automating Suggestions for Well-Being, in: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, pp. 62-68, 2019.

[26] T. Boudaa, M. El Marouani, N. Enneya, Alignment based approach for Arabic textual entailment, Publisher, City, 2019.

[27] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, C. Wang, Machine learning and deep learning methods for cybersecurity, Publisher, City, 2018.