HOW CAN TRANSFORMER-BASED BIDIRECTIONAL ENCODERS ENHANCE THE CLASSIFICATION OF STUDENT PUBLICATIONS ON SOCIAL MEDIA?

OMAR ZAHOUR[1], EL HABIB BENLAHMAR[2], MEHDI EL AMINE [3]

[1]Laboratory of Information Technology and Modelling, Hassan II University of Casablanca, Faculty of Sciences Ben M'SIK, Morocco

[2]Laboratory of Information Technology and Modelling, Hassan II University of Casablanca, Faculty of Sciences Ben M'SIK, Morocco

[3]Laboratory of Information Technology and Modelling, Hassan II University of Casablanca, Faculty of Sciences Ben M'SIK, Morocco

E-mail : [1]orzahour@gmail.com, [2]h.benlahmer@gmail.com , [3]mehdi.elamine-etu@etu.univh2c.ma

ABSTRACT

Text categorization, especially concerning student submissions, constitutes a fundamental undertaking in the realm of natural language processing (NLP). In recent times, bidirectional encoder representations from transformers, commonly known as BERT, a renowned model developed by Google, has been at the forefront in yielding groundbreaking outcomes across a plethora of NLP tasks. In the contemporary era, where text forms the bulk of the data available globally, the significance of automated NLP techniques cannot be understated, emerging not just as a valuable tool but a critical necessity in the sphere of artificial intelligence. While models like BERT have made notable strides, offering unprecedented results in comparison to earlier methods, they sometimes overlook the nuanced local information embedded within the textual data, such as inter-sentential relations. In this study, we delve into the intricacies of leveraging BERT for multi-class text classification, focusing specifically on categorizing student articles pertinent to the educational and vocational guidance sector, in alignment with Holland's RIASEC theory. Our constructed model is adept at determining the precise category of the input publication, operating within a framework that recognizes six distinct classes encompassing our dataset. Conducted through comprehensive experiments utilizing Python, the proposed model exhibits a promising performance benchmark, standing as a potent contender in the ever-evolving domain of text classification and NLP.

Keywords: *Text Categorization; Bidirectional Encoder Representations from Transformers (BERT); Natural Language Processing (NLP); Multi-class Classification; Educational and Vocational Guidance; RIASEC Typology of Holland*

1. INTRODUCTION

Text categorization is a core activity within the domain of natural language processing (NLP), tasked with allocating various textual segments to pertinent categories. Applications of text classification span across several areas, encompassing sentiment analysis [1], question categorization [2], and topical classification [3]. In contemporary settings, deep learning techniques have emerged as standard approaches for conducting text categorization, utilizing frameworks such as convolutional neural networks (CNN) [4], recurrent neural networks (RNN) [5], among other intricate methodologies.

The deep learning approach to text classification involves feeding textual data into a deep neural network to derive a representation of the text. Subsequently, this text representation is input into a softmax function, which calculates the probability distribution across various categories, facilitating the assignment of the text to the most likely category.

Models based on Convolutional Neural Networks (CNN) [4] [6] [7] are adept at capturing local information present in text representations, whereas Recurrent Neural Network (RNN)-based models [8] [9] are proficient in encapsulating long-term sequential information within the textual data. Consequently, to leverage the strengths of both CNN and RNN architectures, various methods have been

Journal of Theoretical and Applied Information Technology
30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

proposed that integrate these two types of networks, including approaches such as C-LSTM [10], CNN-LSTM [11], and DRNN [12]. These hybrid models aim to provide a more comprehensive understanding of text data by assimilating both local and sequential information in the analysis process.

Building upon this, several models incorporate attention mechanisms to enable the model to concentrate on critical information embedded within the text. For instance, the Hierarchical Attention Network (HAN) model [13] employs a tiered attention mechanism to partition the text into two levels: sentences and words, utilizing a bidirectional RNN as an encoder to facilitate nuanced analysis. Meanwhile, the Deep Complex Convolutional Neural Network (DCCNN) [14] initially leverages a multilayer CNN to seize representations of different attributes characterized by n-grams. Following this, it engages the attention mechanism to derive representations by pinpointing and accentuating the most salient characteristics, thereby enhancing the quality of text analysis by focusing on pivotal components within the data.

The MEAN model [15] endeavors to mitigate existing issues by infusing three varieties of emotional linguistic insights into the deep neural network, facilitated by the utilization of attention mechanisms. On the other hand, DiSAN [16] pioneers a novel attention mechanism where the attention interaction between elements of input sequences is both directional and multidimensional, offering a nuanced approach to data interpretation.

Furthermore, there are models where the attention mechanism is not just auxiliary but serves as the core component of the analysis. A case in point is the Bi-BloSAN [17], which introduces the Block Self-Attention mechanism as a primary encoder for text analysis, while concurrently utilizing gated network structures to extract essential features effectively. These advanced models leverage the attention mechanism fundamentally to pinpoint and emphasize more critical characteristics within the data, aligning more closely with human observational patterns compared to traditional approaches like max pooling and average pooling, which may sometimes overlook nuanced details present in the data.

Furthermore, pre-training in linguistic models has demonstrated substantial efficacy in acquiring universal language representations by leveraging extensive volumes of unlabeled data. Notable instances of such pioneering approaches are found in Elmo [18], GPT [19], ULMFiT [20], and BERT [21], which stand as some of the quintessential embodiments of this methodology. These entities represent neural network language models developed from text data through unsupervised learning objectives.

Taking BERT as an instance, it is structured upon a bidirectional multi-layer transformer and undergoes training in raw text for objectives such as masked word prediction and subsequent sentence prediction tasks. To adapt a pre-trained model to specialized tasks, it is imperative to fine-tune them utilizing task-centric training datasets and incorporating additional layers specifically designed for the task at hand post the pre-training phase.

To illustrate, in the context of text categorization endeavors, BERT incorporates a simplistic softmax layer succeeding the pre-trained module. This addition facilitates the refinement process, enabling the creation of advanced models adept at handling text classification tasks for particular datasets with heightened efficiency and accuracy.

The BERT model exhibits remarkable efficacy in text classification assignments owing to its adept language understanding capacities. In addressing the complex issue of categorizing student submissions pertaining to educational and vocational guidance, we leveraged the capabilities of the BERT model. This approach aims to facilitate the automated categorization of these submissions into six distinct classes, drawing inspiration from Holland's renowned model and theory and its corresponding RIASEC typology.

In this research piece, we introduce a tailored BERT model specifically engineered for text classification tasks. This iteration of the model is capable of discerning the category of a given publication with heightened precision and effectiveness, thus serving as a potent tool in streamlining the classification process in line with Holland's established categorizations.

In the rapidly advancing realm of Natural Language Processing (NLP), we take a leap by introducing a bespoke BERT model, fine-tuned specifically for text classification tasks. While BERT in its original form has displayed its prowess across diverse NLP tasks, our tailored version brings in a nuanced layer of optimization. The primary objective of this iteration is to offer an enhanced capability in discerning the category of a given publication. What makes our model unique is its alignment with Holland's renowned categorizations, ensuring that the classification process isn't just technologically advanced, but also theoretically

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sound. Through this combination, we present a tool that boasts heightened precision, streamlined effectiveness, and contextual relevance.

While models like BERT have revolutionized text classification, a major challenge remains: how can we enhance the efficacy of these models to capture nuanced local information, especially when classifying student articles in alignment with Holland's RIASEC theory, while still maintaining top-tier performance?

This article unfolds in the following structure:

Initially, we immerse ourselves in the realm of related work, laying a firm foundation upon which the subsequent sections are built. The ensuing segment is earmarked for a comprehensive exploration of the employed methodology, setting the stage for a deep-dive into the experimental phases and the results gleaned, which constitute the third segment of this article. Bringing our discourse to a close, the final segment encapsulates our conclusions along with potential avenues for future research, offering a glimpse into prospective explorations in this domain.

2. RELATED WORKS

2.1 Deep neural networks

In recent times, deep neural networks have carved a niche for themselves by demonstrating promising results in the domain of natural language processing. Among them, Recurrent Neural Networks (RNN), inclusive of mechanisms like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have emerged as fitting solutions for managing sequences of words. Several nuanced variations have also surfaced, including TreeLSTM [8] and TG-LSTM [9], which bring additional facets to the table.

Parallelly, Convolutional Neural Networks (CNN) have established themselves as a favorite in the deep learning spectrum. VDCNN [6] is pioneering efforts to craft deeper CNN structures specifically targeting text classification. Simultaneously, research [7] has ventured into utilizing an array of filters with varied window dimensions, aiming to extract convolutional features at multiple scales conducive to text classification.

Adding to the innovations, DCNN [22] introduces a dynamic k-max pooling mechanism, while DPCNN [23] is geared towards deepening CNN architectures without substantially escalating computational demands. Further, study [24] unveils a novel approach to weight initialization, enhancing CNN's effectiveness in text classification tasks.

LK-MTL [25], on the other hand, presents a multitasking convolutional neural network equipped with a Leaky unit, introducing memory and a forgetting mechanism to meticulously filter functionality flow between varying tasks. Standing distinct from the mentioned methodologies, charCNN [4] represents a character-level model, proficient in encoding the individual characters present within the input text, adding a unique dimension to the realm of text classification.

Inevitably, several methodologies are in the works to amalgamate the strengths of both CNN and RNN models, aiming to harness a synergized potential. The C-LSTM [10] model takes the initiative in this direction by first employing CNN to seize local textual nuances, followed by leveraging the LSTM network to encode the outcomes derived from each convolutional kernel, thereby grasping global data insights. In a similar vein, the CNN-RNN [11] model adopts a congruent structure, though it distinguishes itself through the unique manner in which the CNN layer interfaces with the RNN layer, optimizing the interaction between these two powerful components.

On another front, the DRNN [12] model ventures to replace the traditional convolutional kernel with an RNN unit. This approach seeks to blend the spatial awareness of the CNN structure with the sequential data handling capabilities of the RNN, fostering a more dynamic and adaptive coding process that effectively captures a wider spectrum of information embedded within the text. This convergence of methodologies hints at a future where hybrid models might reign supreme, offering a nuanced and comprehensive approach to text analysis and classification.

2.2 Pre-training model

In recent times, akin to developments in computer vision research, pretraining models have exhibited exemplary performance across a variety of natural language processing tasks. These models predominantly harness the power of universal language representations, honed through the analysis of vast pools of unlabeled data, followed by the integration of supplementary layers post the preform module to cater to diverse tasks.

Elmo [18], for instance, is committed to extracting context-rich representations from language models. This strategy facilitates groundbreaking accomplishments across several benchmarks in the realm of NLP, notably making

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

strides in question answering [26], sentiment analysis [1], and named entity recognition [27], to name a few. This approach has not only enhanced the efficacy of existing applications but has also paved the way for more nuanced and sophisticated linguistic analyses, underpinning a new wave of innovation in the field of natural language processing.

GPT [19] [33] and ULMFiT [20] engage in pre-training a specified model architecture guided by a language modeling (LM) objective, setting the stage for subsequent fine-tuning to cater to supervised downstream tasks, including but not limited to text classification. Moreover, the noteworthy approach exemplified by the BERT methodology [21] must be highlighted. This method has significantly impacted the field, demonstrating the potency of pre-trained models in facilitating nuanced understanding and categorization of text data, thereby playing a pivotal role in advancing the current landscape of natural language processing tasks. It has paved the way for further research and innovations in the sphere of text analysis and classification, showcasing the utility and versatility of pre-trained models in various complex NLP undertakings.

3. THE PROPOSED METHOD

The BERT algorithm stands firmly on groundbreaking frameworks like the sequence-to-sequence (seq2seq) models and transformer architectures. The seq2seq model is essentially a network structure adept at transmuting a designated sequence of elements (typically words or phrases) into a distinct, yet correlative sequence. This model is proficient at establishing connections between words that appear to hold higher significance or relevance within the context, hence enhancing the analytical depth and context comprehension.

LSTM networks are prime examples of seq2seq models, which demonstrate remarkable proficiency in handling sequential data, where the order of data points is of vital importance. These networks have been particularly efficient in capturing long-term dependencies in sequences, which traditional RNNs struggle with.

On the other hand, the transformer architecture takes on the role of transmuting one sequence into another, distinctly veering away from the reliance on recurrent networks, such as LSTM or GRU. This departure allows for parallel processing of sequences, thereby accelerating the training and prediction processes. Moreover, it incorporates attention mechanisms at its core, facilitating the focus on different parts of the input sequence when producing an output sequence, thereby enabling a deeper understanding and representation of the data. This innovative architecture has been a cornerstone in recent advancements in Natural Language Processing, laying the foundation for models like BERT to excel in various complex tasks and set new benchmarks in performance.

In our classification scheme, we adopted the BERT methodology [21], a potent tool grounded on a bidirectional multi-layer transformer infrastructure. This model is proficiently trained on plain text data, with a primary focus on masked words prediction tasks along with subsequent sentence prediction endeavors. The approach hones in on understanding contextual nuances within textual datasets, thereby fostering an enriched analytical landscape for the classification of various data segments. This strategy facilitates a more nuanced and precise categorization, making it a formidable choice for tackling the complexities inherent in educational and vocational guidance publication classification.

Figure one shows the structure of BERT when it performs text categorization tasks.

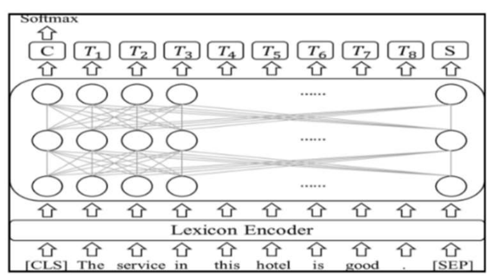

Figure 1: Scheme of model BERT

In the intricate process of language model operations orchestrated by BERT, distinctive markers and encoders play a pivotal role. To delineate the commencement and culmination of a sentence, markers denoted as [CLS] and [SEP] are employed, serving as the start and end flags respectively. These markers not only assist in demarcating individual sentences but also facilitate seamless processing during the tokenization phase, providing structural integrity to the textual data ingested.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

At the succeeding stage, the lexical encoder springs into action, amalgamating a trio of vital incorporations: token incorporations, segmentation incorporations, and position incorporations. This amalgamation process is integral in synthesizing a comprehensive representation of the text input, allowing for a nuanced understanding of the text's structure, semantics, and positional relevance. Furthermore, the sophisticated process aids in facilitating a deeper comprehension of the textual nuances, offering a rich tapestry of information that is ready to be analyzed and processed further.

As we venture deeper into the heart of the BERT model, we encounter the transformative power of the multi-layer self-attention mechanism, housed within the transformer encoder module. This sophisticated mechanism meticulously assigns corresponding output values to each input token, thus effectively capturing the intricate relationships and dependencies between various tokens. The mechanism leverages the power of attention weights to discern the relative importance of each token in the context of others, fostering a robust and highly adaptive model capable of discerning subtle linguistic nuances.

Through this well-coordinated and harmonized process, the BERT model stands as a beacon of innovation in the domain of text classification, offering unparalleled accuracy and efficiency in discerning and categorizing textual data based on Holland's model and theory and its RIASEC typology. This detailed framework, thus, serves as a potent tool in the realm of educational and vocational guidance, ushering in a new era of precision and effectiveness in text classification tasks.

Within the complex structure of the BERT model, several aspects play pivotal roles in achieving refined classification results. One such facet is the generation of a comprehensive representation of the entire body of text, a process characterized by the assimilation of information drawn from every individual word within the text. This representation is significant because it forms a nexus, aggregating insights and nuanced data from various textual elements.

Following this, the text representation is channeled into the Softmax layer, a crucial stage where the classification outcomes are meticulously derived. It is here that the symbol 'C' manifests its significance, attentively discerning the relative importance of each word within the textual corpus. It operates under the principle that each word within the dataset is treated as an equal and independent entity, thus ensuring a level of objectivity in the classification process.

Nevertheless, it is imperative to note a particular limitation inherent in this approach. The model, in its current configuration, overlooks the nuanced information encapsulated within specific fragments or sentences present in the text. This means that while it maintains a holistic approach, focusing on individual words, it potentially misses the deeper contextual richness that certain phrases or sentence structures might offer. This could potentially be seen as a drawback as it doesn't fully capitalize on the complex interrelations and subtleties that are often embedded within sentence structures and fragments, which could have been instrumental in fostering a more nuanced and comprehensive analysis.

In future iterations of this model, incorporating mechanisms to capture these intricate details found in fragments or sentences could further enhance the accuracy and depth of its classification capabilities. By evolving to recognize and utilize these deeper layers of textual information, the model can aspire to achieve an even more sophisticated understanding and interpretation of textual data, paving the way for more nuanced and insightful classifications.

4. EXPERIMENTAL ASSESSMENT AND OUTCOME ANALYSIS

4.1 Dataset and Features

Our dataset is derived from the RIASEC test, which is grounded on Holland's theory [28], [29], [30], [31]. This dataset comprises two primary columns, as delineated below:

Publication: This column encapsulates student submissions that scrutinize and evaluate individual personality traits of the users.

Class: Within this column, we categorize the submissions into six distinctive classes (or labels), namely:

1) Realistic. (0)
2) Investigative. (1)
3) Artistic. (2)

Journal of Theoretical and Applied Information Technology
30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4) Social. (3)
5) Enterprising. (4)
6) Conventional. (5)

4.2 Experiment Steps

For the successful implementation and subsequent inference using the BERT model for text classification, it is vital to align our dataset according to the specific prerequisites dictated by the BERT framework. The BERT model anticipates datasets to adhere to a defined structure characterized by four principal components:

- **Guide:** This component functions as a unique identifier, encapsulating individual instances within the dataset.
- **Text_a:** This segment houses the text material which is designated for classification into predetermined categories.
- **Text_b**: This segment is primarily utilized to discern relationships between various sentences; however, it remains non-essential to our current classification endeavor.
- **Label:** This vital section accommodates the labels or categories which signify the respective classifications of the contents in the 'text_a' section.

In our dataset configuration, we've designated the 'Publication' field to function as 'text_a' and the 'Class' field to serve as the 'label'. Our subsequent step entails transforming this dataset to align with the structure mandated by BERT. In accomplishing this, we will utilize the InputExample class available within the BERT library to create objects embodying each of these attributes for every record in our dataset.

With the dataset now correctly formatted to cater to BERT's specifications, we are prepared to initiate the data preprocessing phase.

Here is our planned approach:

- **Tokenization and Encoding:**

During this stage, we will process the text data encompassed in both the training and validation sets. The primary step is tokenization, a method which segregates text into individual words or subwords. Following tokenization is the encoding phase, where these tokens are transformed into numerical representations, a format that is compatible with the BERT model. It is imperative to note that text sequences will be confined to a maximum of 128 tokens. Sequences exceeding this length will be truncated, while shorter sequences will be supplemented with special tokens to meet this predefined length. Subsequently, this refined data will be transformed into PyTorch tensors to facilitate further operations.

- **Creating Datasets:**

Following the tokenization and encoding processes, we construct datasets by amalgamating the encoded textual data with their respective labels. This structure meticulously aligns each input text sequence with its corresponding label, thereby facilitating the model to accurately discern the respective category or class each text is affiliated with during the training and validation phases

- **Data Loaders:**

DataLoaders are devised to proficiently handle the datasets. These tools segregate the data into batches, with each batch encapsulating 16 text sequences along with their corresponding labels, as delineated. Throughout the training regimen, the sequence of data is randomized in each epoch to circumvent any potential learning of patterns based on sequence order, thereby bolstering the efficacy of training. By furnishing batches of data to the model in each iteration, DataLoaders facilitate a streamlined and efficient training process.

After the data preprocessing steps, the code proceeds to train a BERT-based model for text classification. It does this by repeatedly exposing the model to the training data in a series of iterations known as epochs. During each epoch, the model learns from the training data, adjusting its internal parameters to better understand and classify text into predefined categories. The optimizer is responsible for updating these parameters based on how well the model is performing, as measured by a loss function.

After each epoch of training, the model's performance is evaluated on a separate validation dataset to assess how well it generalizes to new, unseen data. The goal is to find the best set of model parameters that result in accurate predictions on the validation data. Once the desired number of epochs is completed, the trained model is saved for later use or deployment, enabling it to classify text into predefined categories effectively.

Journal of Theoretical and Applied Information Technology
30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.3 Results

We used the BERT algorithm to classify the category of student publications according to Holland's RIASEC typology.

These figures show the results obtained by this model:

```
Epoch 1: Accuracy = 0.85273
Epoch 2: Accuracy = 0.92874
Epoch 3: Accuracy = 0.94299
Epoch 4: Accuracy = 0.95012
Epoch 5: Accuracy = 0.94537
```

Figure 2: The Accuracy of the model

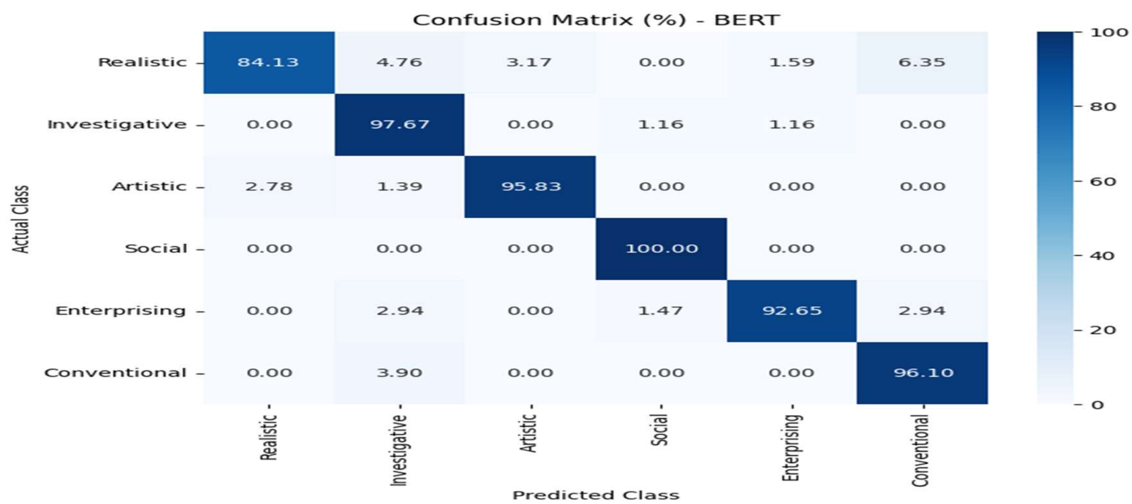

Figure 3: Confusion matrix of BERT model

```
Text: I enjoy working with my hands and solving practical problems. Building and fixing things is a natural fit for me.
Predicted Label: Realistic

Text: I'm naturally curious and love exploring complex ideas. Research and analysis are my strengths.
Predicted Label: Investigative

Text: Creativity flows through me, and I find joy in expressing myself through various art forms like painting, writing, and music.
Predicted Label: Artistic

Text: Helping and connecting with others is where I thrive. I'm drawn to roles that involve teaching, counseling, or assisting people in need.
Predicted Label: Social

Text: I have a strong drive for leadership and enjoy taking risks. Entrepreneurship and sales allow me to use my persuasive skills.
Predicted Label: Enterprising

Text: I'm detail-oriented and excel in organized environments. I find satisfaction in tasks that require precision, like data analysis or accounting.
Predicted Label: Conventional
```

Figure 4: Test of our classification model

5. CONLUSION

This study delved into the utilization of BERT, a potent NLP algorithm, for the classification of student publications within the framework of the RIASEC model. Our journey commenced with the meticulous preprocessing of raw text to align it with BERT's format [33], thereby enabling its contextual comprehension.

As the field of "Student's Orientation" continues to advance, it opens up several promising avenues for future research. One such direction is the expansion of the scope of student publications, moving beyond traditional academic texts to include digital content, social media contributions, and portfolio materials.

Exploring interdisciplinary approaches that incorporate behavioral science, psychology, and sociology can offer valuable insights into the intricate interplay between student publications and career interests.

Conducting research to examine the influence of career-oriented guidance on students' academic performance and professional development holds the potential to shape personalized learning and career planning strategies. Recent studies have also ventured into extending and adapting the RIASEC model to enhance its relevance and applicability.

Certain researchers have embraced holistic and dynamic approaches to career guidance, considering factors like individual aspirations, personal values, socioemotional elements, and job satisfaction. Additionally, there is research exploring the connections between RIASEC personality types and other personality models, such as the Big Five, in order to gain deeper insights into the interplay between personality traits and career preferences.

Guiding the youth, a pivotal concern in today's context, goes beyond simply making academic or professional choices. In an era flooded with career and academic paths, young individuals often find themselves at crossroads, unsure of which direction to take. This guidance isn't just about picking a career or an academic stream. It encompasses understanding oneself, discovering innate passions and talents, and aligning these with real-world opportunities. Moreover, with the rising challenges our society faces, such as rapid technological shifts and economic uncertainties, there's an imperative to offer the youth tailored advice and support. This ensures they navigate this ever-changing landscape successfully, blossoming into fulfilled and contributive citizens [32].

In summation, the adoption of BERT for the categorization of student articles as per Holland's RIASEC theory heralds a promising direction for NLP applications in educational sectors. The findings advocate for the broader implementation of such advanced models, given their proven efficiency and the depth of understanding they offer. Future endeavors could further explore model hybridizations or modifications to accentuate the strengths and mitigate potential weaknesses observed.

REFERENCES:

[1] Maas, A.L., et al.:" Learning word vectors for sentiment analysis. "In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1. Association for Computational Linguistics (2011).

[2] Zhang, D., Lee, W.S.: 'Question classification using support vector machines.' In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM (2003).

[3] Wang, S., Manning, C.D.: ‘Baselines and bigrams: simple, good sentiment and topic classification.’ In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, vol. 2. Association for Computational Linguistics (2012).

[4] Zhang, X., Zhao, J., LeCun, Y.: ‘Character-level convolutional networks for text classification.’ In: Advances in Neural Information Processing Systems (2015).

[5] Chung, J., et al. ‘Empirical evaluation of gated recurrent neural networks on sequence modeling.’ arXiv preprint arXiv:1412.3555 (2014).

[6] Conneau, A., et al." Very deep convolutional networks for text classification." arXiv preprint arXiv:1606.01781 (2016).

[7] Kim, Y."Convolutional neural networks for sentence classification. ",arXiv preprint arXiv:1408.5882 (2014).

[8] Tai, K.S., Socher, R., Manning, C.D."Improved semantic representations from tree-structured long short-term memory networks.", arXiv preprint arXiv:1503.00075 (2015).

[9] Huang, M., Qian, Q., Zhu, X." Encoding syntactic knowledge in neural networks for sentiment classification. ",ACM Trans. Inf. Syst. (TOIS) 35(3), 26 (2017).

[10] Zhou, C., et al." A-C-LSTM neural network for text classification.", arXiv preprint arXiv:1511.08630 (2015).

[11] Xiao, Y., Cho, K." Efficient character-level document classification by combining convolution and recurrent layers.", arXiv preprint arXiv:1602.00367 (2016).

[12] Wang, B." Disconnected recurrent neural networks for text categorization.", In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Long Papers, vol. 1 (2018).

[13] Yang, Z., et al."Hierarchical attention networks for document classification. ", In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (2016).

[14] Wang, S., Huang, M., Deng, Z."Densely connected CNN with multi-scale feature attention for text classification. ", In IJCAI (2018).

[15] Lei, Z., et al."A multi-sentiment-resource enhanced attention network for sentiment classification. ",arXiv preprint arXiv:1807.04990 (2018).

[16] Shen, T., et al." DiSAN: directional self-attention network for RNN/CNNfree language understanding. ", In Thirty-Second AAAI Conference on Artificial Intelligence (2018).

[17] Shen, T., et al."Bi-directional block self-attention for fast and memoryefficient sequence modeling. ",arXiv preprint arXiv:1804.00857 (2018).

[18] Peters, M.E., et al."Deep contextualized word representations.", arXiv preprint arXiv:1802.05365 (2018).

[19] Radford, A., et al." Improving language understanding by generative pretraining", (2018). https://s3-us-west-2.amazonaws.com/openaiassets/research-covers/langu ageunsupervised/languageunderstandingpaper.pdf.

[20] Howard, J., Ruder, S." Universal language model fine-tuning for text classification.", arXiv preprint arXiv:1801.06146 (2018).

[21] Devlin, J., et al." Bert: pre-training of deep bidirectional transformers for language understanding.", arXiv preprint arXiv:1810.04805 (2018).

[22] Kalchbrenner, N., Grefenstette, E., Blunsom, P."A convolutional neural network for modelling sentences.", arXiv preprint arXiv:1404.2188 (2014).

[23] Johnson, R., Zhang, T."Deep pyramid convolutional neural networks for text categorization.", In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Long Papers, vol. 1 (2017).

[24] Li, S., et al." Initializing convolutional filters with semantic features for text classification.", In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (2017).

[25] Xiao, L., et al."Learning what to share: leaky multi-task network for text classification.", In: Proceedings of the 27th International Conference on Computational Linguistics (2018).

[26] Rajpurkar, P., et al." SQuAD: 100,000+ questions for machine comprehension of text. ",arXiv preprint arXiv:1606.05250 (2016).

[27] Sang, E.F., De Meulder, F." Introduction to the CoNLL-2003 shared task: languageindependent named entity recognition.", arXiv preprint cs/0306050 (2003).

[28] Omar Zahour, El Habib Benlahmar, Ahmed Eddaoui " E-orientation : entre prescription des théories et prise de décision " Conference TIM'16,2016.

[29] Omar Zahour, El Habib Benlahmar,Ahmed Eddaoui " E-Orientation: Between prescription of theories and decision-making " Conference SITA'16, October 2016.

[30] Omar Zahour, El Habib Benlahmar, Ahmed Eddaoui "E-orientation : Vers une modélisation des facteurs d'orientation scolaire " Conference TIM'18, 2018.

[31] Omar Zahour, El Habib Benlahmar, Ahmed Eddaoui, Oumaima Hourrane "Automatic Classification of Academic and Vocational Guidance Questions using Multiclass Neural Network January 2019, International Journal of Advanced Computer Science and Applications 10(10); DOI: 10.14569/IJACSA.2019.0101072. (2019)

[32] Ihya, Rachida & Namir, Abdelwahed & Filali, Sanaa & Guerss, Fatima zahra & Haddani, Hajar & Aitdaoud, Mohammed. ACCEPTANCE MODEL PREDICTION'S FOR E-ORIENTATION SYSTEMS CASE OF STUDY: PLATFORM "orientation-chabab.com". Journal of Theoretical and Applied Information Technology. 97. 4227-4238. (2019).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[33] Lund, Brady D., and Ting Wang. "Chatting about ChatGPT: how may AI and GPT impact academia and libraries?" Library Hi Tech News 40.3 (2023): 26-29.

[34] Alammary, Ali Saleh. 2022. "BERT Models for Arabic Text Classification: A Systematic Review" *Applied Sciences* 12, no. 11: 5720. https://doi.org/10.3390/app12115720. (2022).