




Article

CVs Classification Using Neural Network Approaches Combined with BERT and Gensim: CVs of Moroccan Engineering Students

Aniss Qostal ^{1,*} , Aniss Moumen ²  and Younes Lakhri ¹ 

¹ Intelligent Systems, Georesources and Renewable Energies Laboratory (SIGER IN FRENCH), Sidi Mohamed Ben Abdellah University, FST, Fez 30050, Morocco; younes.lakhri@usmba.ac.ma

² Laboratory of Engineering Sciences, National School of Applied Sciences, Ibn Tofail University, Kenitra 14000, Morocco; aniss.oumoumen@uit.ac.ma

* Correspondence: aniss.qostal@usmba.ac.ma

Abstract: Deep learning (DL)-oriented document processing is widely used in different fields for extraction, recognition, and classification processes from raw corpus of data. The article examines the application of deep learning approaches, based on different neural network methods, including Gated Recurrent Unit (GRU), long short-term memory (LSTM), and convolutional neural networks (CNNs). The compared models were combined with two different word embedding techniques, namely: Bidirectional Encoder Representations from Transformers (BERT) and Gensim Word2Vec. The models are designed to evaluate the performance of architectures based on neural network techniques for the classification of CVs of Moroccan engineering students at ENSAK (National School of Applied Sciences of Kenitra, Ibn Tofail University). The used dataset included CVs collected from engineering students at ENSAK in 2023 for a project on the employability of Moroccan engineers in which new approaches were applied, especially machine learning, deep learning, and big data. Accordingly, 867 resumes were collected from five specialties of study (Electrical Engineering (ELE), Networks and Systems Telecommunications (NST), Computer Engineering (CE), Automotive Mechatronics Engineering (AutoMec), Industrial Engineering (Indus)). The results showed that the proposed models based on the BERT embedding approach had more accuracy compared to models based on the Gensim Word2Vec embedding approach. Accordingly, the CNN-GRU/BERT model achieved slightly better accuracy with 0.9351 compared to other hybrid models. On the other hand, single learning models also have good metrics, especially based on BERT embedding architectures, where CNN has the best accuracy with 0.9188.

Keywords: Gated Recurrent Unit (GRU); long short-term memory (LSTM); convolutional neural network (CNN); BERT; Gensim; Moroccan engineering students; Ibn Tofail University; CVs; ENSAK



Citation: Qostal, A.; Moumen, A.; Lakhri, Y. CVs Classification Using Neural Network Approaches Combined with BERT and Gensim: CVs of Moroccan Engineering Students. *Data* **2024**, *9*, 74. <https://doi.org/10.3390/data9060074>

Academic Editors: Kassim S. Mwitondi and Joaquín Torres-Sospedra

Received: 27 February 2024

Revised: 7 May 2024

Accepted: 17 May 2024

Published: 24 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of Artificial Intelligence (AI) including deep learning (DL) and machine learning (ML) approaches to process large corpus, with the main objectives of information extraction, discovery of hidden patterns, as well as classification and categorization, has gained popularity in certain fields [1]. Therefore, in many areas of the healthcare sector, clinical decision-making based on scanned images and laboratory samples has produced good results in classification, interpretation, and even prediction for patients [2,3]. Also, several approaches based on machine learning have been proposed for evaluating the performance in different fields such as the educational systems and assessing student performance [4,5], the sentimental analysis based on machine learning, which can be applied in a variety of fields, such as politics [6], and tourism [7].

Therefore, the use of machine learning can be an innovative approach to improve the employability of students, support their career prospects, and identify their skills, and

weaknesses based on labor market requirements [8]. Accordingly, different classification approaches can be applied to capture relevant information and discover hidden patterns by analyzing data related to students' activities [9]. Furthermore, a variety of data sources were used, including professional social networks (e.g., LinkedIn platform), CV and resume platforms, university student profiles, and even raw digital documents [10,11]. Thus, even though CVs are considered unstructured, and noisy documents, student CVs remain important documents that contain relevant professional and personal information and are directly oriented towards employability [12]. Accordingly, we propose our research on creating an employability model for Moroccan engineering students using combined machine learning to extract relevant information from their CVs.

Thus, when processing data from previous sources to obtain targeted results, a supervised or unsupervised machine learning approach can be used, including classification and clustering algorithms and even neural network approaches where data can be with or without labels [13,14]. Recently, the use of models based on deep learning techniques namely the neural networks approaches such as gated recurrent unit (GRU), long short-term memory (LSTM), and convolutional nNeuralnetwork (CNN) on text classification are generally chosen for their good performance on sequential tasks such as NLP and text classification [15,16].

In our case, the data under analysis consists of CVs of students containing a variety of information including personal and professional information, hard skills, soft skills, and academic background. Therefore, the use of deep learning techniques in text classification showed good performance and accuracy with new techniques based on neural network approaches [15].

Accordingly, for the classification of ENSAK's engineering students' CVs, we propose two different models combining two embedding techniques namely: BERT and Gensim with three approaches of neural network architectures namely: GRU, LSTM, and CNN. The results obtained from this study can serve as a basis for an in-depth analysis of the labor market and skills of Moroccan engineers. Additionally, the application of deep learning could contribute to the development of recommendation systems based on unstructured data associated with candidates and job offers intending to reduce unemployment.

Previously, different approaches have been proposed to use data such as job postings as a basis for building a recommendation system based on web-mined job postings and job seekers' skills, including classification and recommendation for job seekers and job providers [17]. Otherwise, the use student documents, notably CVs. and classification resumes are still little explored, particularly for the Moroccan context, and using new approaches such as BERT and Neural Network architectures.

The paper is organized as follows; the second section presents the related works that have been applied to identify relevant information based on intelligent approaches and machine learning models especially for the employability on the Moroccan context. The third section presented our methodology adopted in this paper, the data collection method and machine learning model used for both architectures (BERT and Gensim). Finally, the fourth section details the findings and the interpretation of the results obtained based on the different approaches adopted.

2. Related Works

The application of machine learning and deep learning in subjects related to students and graduates is no exception; since the application of those new approaches has shown good results in helping and promoting decision-making in the education system [4]. Therefore, these models are based on data from various sources and systems related to the academic process, including Learning Management Systems (LMS), MOOCs, Student Information Systems (SIS), and Intelligent Teaching Systems (ITS) [4]. Accordingly, DL and NLP have been introduced into the education system in various areas, such as interpreting student behavior, detecting a lack of student motivation, analyzing the level of interest in lessons, predicting academic results, and even preventing school dropout problems [5,18].

Several studies have been carried out in the Moroccan educational context, where [19] proposed a model with an accuracy of 71% based on data from the Scholar Management System MASSAR and targeted baccalaureate students. Accordingly, based on the same experience, the prediction of accomplishment is very important for this group of students, where the decision to support or reinforce courses is suitable for both students and the educational system [19]. Furthermore, the author [20] has focused his research on using massive open online course (MOOCS) data to classify students and predict their dropout problems, therefore, the objective is a predictive model where the issue of huge dropout rate reaches 90%. Based on machine learning, the accuracy of the compared models was classified as fellow Support Vector Machines (85.2%), K Nearest Neighbors (83.9%), Decision Trees (77%), Naive Bayes (85.5%), and Logistic Regressions (86.8%) with a combinatorial approach based on voting (92%) [20].

In addition, the author [21] conducted research into the prediction of on-time graduation rates among Moroccan students using models utilizing Support Vector Machines, Decision Trees, Naive Bayes, Logistic Regression, and Random forest. As a result, Random Forest's accuracy of 77% was found to be the highest among the models when analyzing academic success factors in Moroccan universities. Moreover, Ouatik [22] proposed a model based on a Big Data architecture (Hadoop and MapReduce) using neural networks, naive Bayes, and K-nearest neighbors to classify student orientation. Based on the results, Naive Bayes is more accurate and efficient when processing on-time data processing with an accuracy reaching 96%. Further, based on a systematic review of the models proposed for the use of Big Data and machine learning for employability, the author [23] proposed an intelligent system for employability based on Big Data and machine learning in Moroccan contexts using different data sources.

Meanwhile, machine learning and deep learning have also gained a great interest after the graduation of students, especially when it comes to predicting their employability for different stakeholders including the education system, employers, and graduates [24,25]. To build their models, different factors are taken into account, including technical skills, soft skills, personality traits, demographics, extracurricular activities, and internships [26]. Hence, different sources are considered, including recruitment platforms, professional social media platforms, CV platforms, and unstructured CV files [26]. Therefore, the targeted results can be binary results such as employability: [Employable or unemployable], or even multi-classification models that correspond to the labor market situation according to different approaches (SVM, ANN, LR, AdaBoost, etc.) [26].

The first approach involves extracting relevant information from different corpus and detecting features by using Named Entity Recognition (NER) and Named Entity Normalization (NEN) [27]. Accordingly, resumes and CVs classification are widely used based on DL, where the hiring process involves the analysis of each candidate's documents which are usually unstructured documents with very noisy data. The process of extracting entities like different attributes and skills is considered to be a challenging task for recruiters, hiring managers, and intelligent models [28]. Therefore, those approaches aim to develop an automated model that can extract relevant information such as skills and personal characteristics for job matching and recommendation (Table 1). For the NER approach, many researchers have used a controlled dictionary called Folksonomy and Taxonomy, such as ESCO (European) and O*NET (American), where the European and American skills, competencies, qualifications, and professions classification were applied [29,30].

Table 1. Predicting employability models using different approaches of machine learning.

Study	Context	Model	Accuracy
[24]	Dataset from the career center of technological institute of the Philippines, Manila with 27,000 information of students with 3000 observations and 9 features of each student	- Decision Tree (DT) - Random Forest (RF) - Support Vector Machine (SVM)	- DT: 85% - RF: 84% - SVM: 92.22%
[31]	Dataset based on mock job interview results with three thousand (3000) observations and twelve (12) features, student performance rating of the on-the-job training students collected	- Decision Tree (DT) - Random Forest (RF) - Support Vector Machine (SVM) - K-Nearest Neighbor (KNN) - Logistic Regression (LR)	- DT: 56.36% - RF: 64% - SVM: 91.22% - KNN: 79.99% - LR: 65.45%
[26]	Dataset (296 records) from survey of graduates and employers in Egypt of training skills, soft skills, and hard skills	- Decision Tree (DT) - Gaussian NB - Logistic Regression (LR) - Random Forest (RF) - Support Vector Machine (SVM)	- DT: 100% - NB: 92% - LR: 98% - RF: 97% - SVM: 98%
[32]	Proposed models for predicted performance and students' employability. Primary datasets of 218 graduate students of higher educational institutions (heis).	- Random Forest RF - Logistic Regression LR - Decision Tree	- RF: 98% - LR: 94% - DT: 97%
[33]	Hybrid DNN model for predicting students' employability using a machine learning approach	- Random Forest (RF) - Support Vector Machine (SVM)	- SVM: 92% - RF: 89% - GB: 84% - XGB: 81% - LR: 86%

However, in the Moroccan context, the application of machine learning to employability prediction is still in its infancy. Nevertheless, some research has given a roadmap in this direction, by analyzing the skills needed within the local professional market, extracting different features and classifications. where the experience led by [34] through neural network approach LSTM of both demands and offers and using word embedding vector supplemented respectively to the taxonomy ESCO database has shown relevant classification for different decision makers based on the priority of different features (explicit skills, soft skills, demographic and geographic information, experiences, etc.).

On the other hand, ref. [35] used combined Word embedding techniques word2Vec and neural network DNN approach for the classification of IT resumes (Web/software development, Network engineering, Embedded software engineering, Testing engineering, Business intelligence, Big data development, Data science, Information systems management, Database administration) and labor market in the Moroccan context. In addition, those approaches can be applied to a recommendation system, where [36] proposed a model as a Recommendation system using data from Moroccan E-recruitment platforms (Rekrute, Emploi.ma, LinkedIn, etc.) and based on a Classification approach including Weighted Semantic Network and Vector Space Model (VSM).

3. Materials and Methods

3.1. Sampling and Data Collection Method

The dataset of CVs was collected as part and continuation of a previous study carried out on the students' skills and their perceptions of labor market demands [37], which aims to uncover students' perceptions of their market-ready skills, including their assessment of the importance of each skill, such as technical and soft skills. On the other hand, the present study aims to use CVs collected from the same sample of ENSAK students (Table 2) using new approaches presented on machine learning to examine the performance of each model for automated and intelligent for classification of their CVs. Therefore, the ENSA network (Ecole Nationales des Sciences Appliquées) is the largest engineering school in Morocco, with eleven establishments in the main regions. The network has thousands of students in various modern fields of study relevant to the professional workplace. The survey was conducted over a six-month period, between January 2023 and June 2023; the students' responses were collected using the Google Forms application with consideration of the confidentiality of personal information in the data collected.

Table 2. Information and characteristics of respondents (sample = 867 students).

ENSA Kenitra	Department	Total
Department	Computer Engineering (CE)	263
	Networks and Systems Telecommunications (NST)	134
	Automotive Mechatronics Engineering (AutoMec)	149
	Industrial Engineering (Indus)	114
	Electrical Engineering (ELE)	207
Total		867

The dataset contained 867 CVs from five departments (Electrical Engineering (ELE), Networks and Systems Telecommunications (NST), Computer Engineering (CE), Automotive Mechatronics Engineering (AutoMec), and Industrial Engineering (Indus). Accordingly, the main objective of the current research is the study the relevance of models built using deep learning, namely, neural network approaches and word embedding techniques to the topic related the student employability using information extracted from their CVs.

3.2. Experiment and Problem Definition

In the first step, the experiment started by loading the collected dataset, which was composed of unstructured files of different formats (Table 3). Next, we generated a global CSV file by extracting text from each document where the data were labeled according to the specialty of the study. Therefore, after loading the final CSV file and removing the noisy data (symbols, punctuation, numbers, etc.) as outlined in the preprocessing and cleaning sections, the generated dataframe can be used for the experiment. As a result, the data were loaded into different models for the generation of the vector representation of the text based on the BERT and Gensim embedding techniques. Finally, the output of each embedding technique was loaded as input for the different neural networks adopted in the research including GRU, LSTM, and CNN.

Table 3. File types that make up the dataset of student resumes.

Type File	Total
Docx	321
Pdf	308
Png	123
Jpg/Jpeg	115

3.3. Architecture of the Proposed Solution

The proposed models considered the set CVs as the input $C = \{c_1, \dots, c_n\}$ and the student's specialty $S: \{s_1, \dots, s_k\}$ as the targets. The models classified and measured the probabilities $\{(c_1, s_1), \dots, (c_n, s_k)\}$. The inputs to the classifier are the training data, which are a finite sequence of $S \times C$ pairs. The output of the classifier is the function $f: C \rightarrow S$ that predicts $s \in S$ for new samples in C (Figure 1). The basic basis of the models is to construct neural network architectures based on interconnected neurons in different layers, simulating the human neural brain metaphor for information processing [38]. Therefore, after the common step of preprocessing the text, the first method based on the embedding layer is generated based on Gensim tokenization, and the target output (Specialty of the student) is presented on one hot encoding. For the second models that were based on the BERT architecture, the input comes from the embedding layer where the text has gone through the steps of applying certain features and tokenization respecting pre-trained models and generating adequate tokens such as [ids_token], [mask_token], and [padd_token] while adding special tokens including [CLS] and [PAD]. The text classification system can be decomposed into the following four stages: feature extraction, dimension reductions, classifier selection, and evaluation.

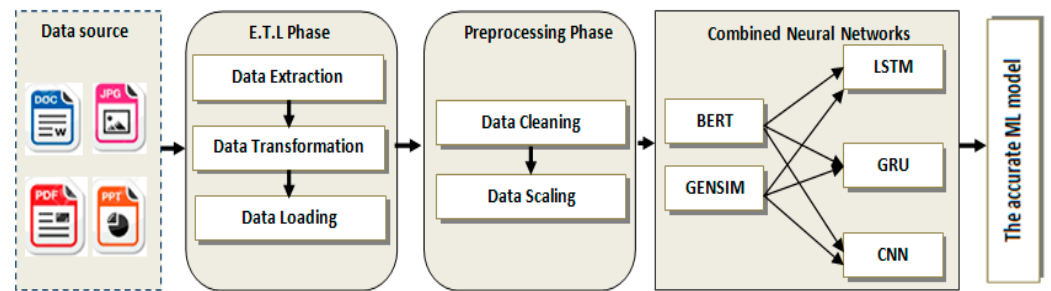


Figure 1. Combined neural network models with Bert/Gensim

3.4. Methods

3.4.1. Long Short-Term Memory (LSTM)

Long short-term memory (LSTM) is considered a standard RNN and was introduced in 1997 [39]. This model has proven successful in solving the vanishing error problem that arises in standard RNN models, as it offers the possibility to use a Constant Error Carousel (CEC), where memory cells are used instead of non-linear activation functions such as Tanh or Sigmoid. This solution enables LSTM models to store and transmit information over the long term (Figure 2).

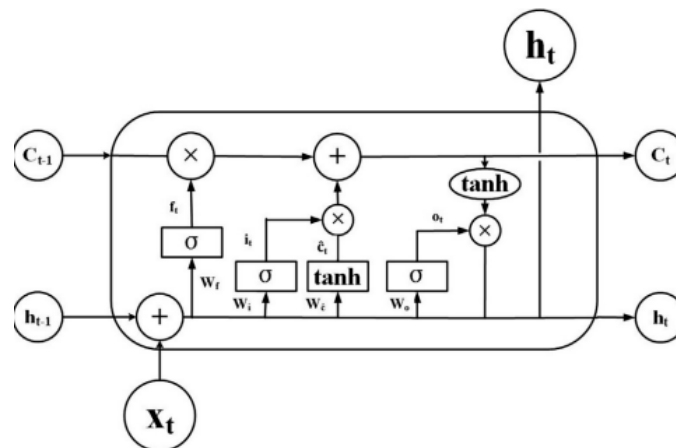


Figure 2. Cell structure and equations describing LSTM gates [40].

Here, σ is the sigmoid function, \tanh is the hyperbolic tangent function, and i , f , o , C , and \hat{C} are the input gate, forget gate, output gate, the content of the memory unit, and the content of the new memory unit, respectively. The sigmoid function is used to form three gates in a memory cell, while the \tanh function is used to improve the performance of a specific memory cell.

3.4.2. Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) was introduced in 2014 as a solution to LSTM's complexity and as a solution to the vanishing gradient problem [41,42]. Moreover, by implementing gating mechanisms within their networks, GRU and LSTM can capture and propagate information over long sequences. Thus, GRU consists of the following components (Figure 3):

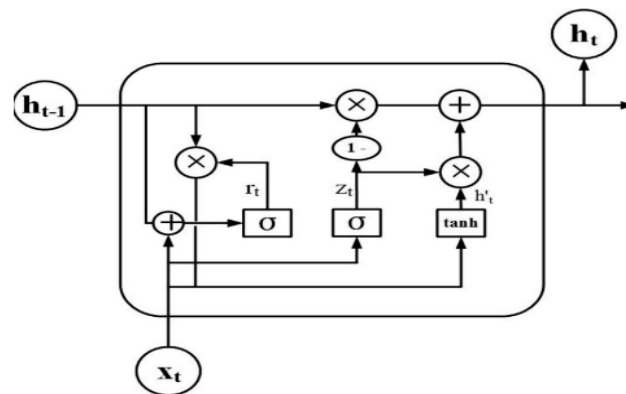


Figure 3. Diagram of a one-unit Gated Recurrent Unit (GRU) [40].

Update Gate (z_t): This calculates how much information from the past should be carried forward.

Reset Gate (r_t): This determines how much information should be forgotten about the past.

Current Memory Content (h_t): This value represents the network's current state.

The update gate and reset gate are responsible for controlling the information flow, and their values are learned adaptively during the training process. The GRU architecture is more computationally efficient than LSTM because it has fewer parameters. However, both GRU and LSTM are widely used in natural language processing, speech recognition, and various sequence modeling tasks.

3.4.3. The Convolutional Neural Network (CNN)

The convolutional neural network (CNN) architecture is a model of deep neural networks used primarily to analyze visual data [43]. CNNs are effectively used in various deep learning methods, including image classification, object detection, text classification, etc. Based on input data, they can learn spatial hierarchies of features automatically and adaptively [43]. Important elements of a CNN include:

Convolutional Layers: These layers operate on the input data using convolutional operations. To detect patterns or features in the input data (image, Text, etc.), convolution applies a filter also called a kernel (Conv1d, Conv2d, and Conv3d). A convolutional layer assists the network in learning how to represent the features hierarchically.

Pooling layer: Convolutional layers are typically followed by a pooling layer. In pooling, the width and height of feature maps are reduced, but their depth is maintained (the number of channels). Pooling operations can be performed in several different ways, including max pooling, which retains the maximum value in a region, and average pooling, which retains an average value

Activation Functions: To introduce non-linearity into the model, non-linear activation functions, such as ReLU (Rectified Linear Unit), Softmax, and Sigmoid, are applied. In this way, the neural network will be able to learn more complex patterns and relationships between the data.

Fully Connected Layers: A convolutional network typically contains several layers of pooling and convolution, followed by a layer of fully connected connections. As a result, the network can make final predictions where each neuron in these layers is connected to the neuron in the previous layer and subsequent layers.

Flattening: Convolutional and pooling outputs are flattened into one-dimensional vectors before the fully connected layers.

3.4.4. The Bidirectional Encoder Representations from Transformers (BERT)

The Bidirectional Encoder Representations from Transformers (BERT) model has been under development since A. Vaswani et al., published “All Attention You Need” in 2017 at Google Labs [44]. The model is an embedding layer of pre-trained bidirectional representations from a large collection of unsupervised text corpus, including Wikipedia and BookCorpus [45]. The related BERT models are BERTBASE and BERTLARGE. Transformer encoding and decoding layers are based on multiple “heads”. The main features and mechanisms of the BERT model are as follows:

Contextual Word Embeddings

Unlike traditional embedding techniques such as Word2Vec and GloVe, BERT produces contextual word embeddings; the weights of words are recorded in the context of a sequence or sentence. The model is based on a bidirectional encoder function, where the input text is entered sequentially (left to right or right to left).

Transformer Architecture:

To analyze a sequential data stream, BERT uses neural network architecture. Two main components are represented in the model, the encoder portion is responsible for the reading of input and the decoder portion responsible for the prediction task; the model is capable of working the two parts simultaneously. BERT, which stands for Bidirectional Encoder Representations from Transformers, is a DL model that can be trained using two methods: Next Sentence Prediction (NSP) and Masked Language Model (MLM). The mechanism used by the BERT model is based on an innovative approach to tokenization methods. BERT uses different and specialized markers during the learning process, such as [CLS] to indicate the beginning of the sequence; [SEP] to indicate the end of the sequence; [PAD] to pad if the sequence lengths are different; and [UNK] to indicate an unknown word in the sequence.

As shown in Figure 4, the first step for building the BERT-based embedding method is to import the required library HuggingFACE and define the pre-trained BERT model to be used. In our case we took different considerations, firstly, the majority of resumes are written in French, and secondly, the average length of the sequence.

The overall steps used in the majority of models of BERT are typically composed of the following steps: tokenization, padding, numericalization, and embedding [45]. For the tokenization, the [ids_token], [mask_token], and other specialized tokens are generated, such as [CLS] mentioning the start of the sequence; [SEP] for the end of the sequence; [PAD] for padding when the sequences do not have the same length; and [UNK] for unknown words in sequences. Since most resumes are in French, we used the un-cased BERT model architecture with 12 layers, 768 hidden nodes, and 12 attention heads with 110M parameters. Token representations are computed by the first-level encoder and then used by the second-level encoder. The whole process is repeated until the 12th encoder is reached, which is the final encoder. Based on the output, the obtained matrix had a size of 256×768 , where 256 represents the number of tokens in the sequence and 768 represents the hidden size.

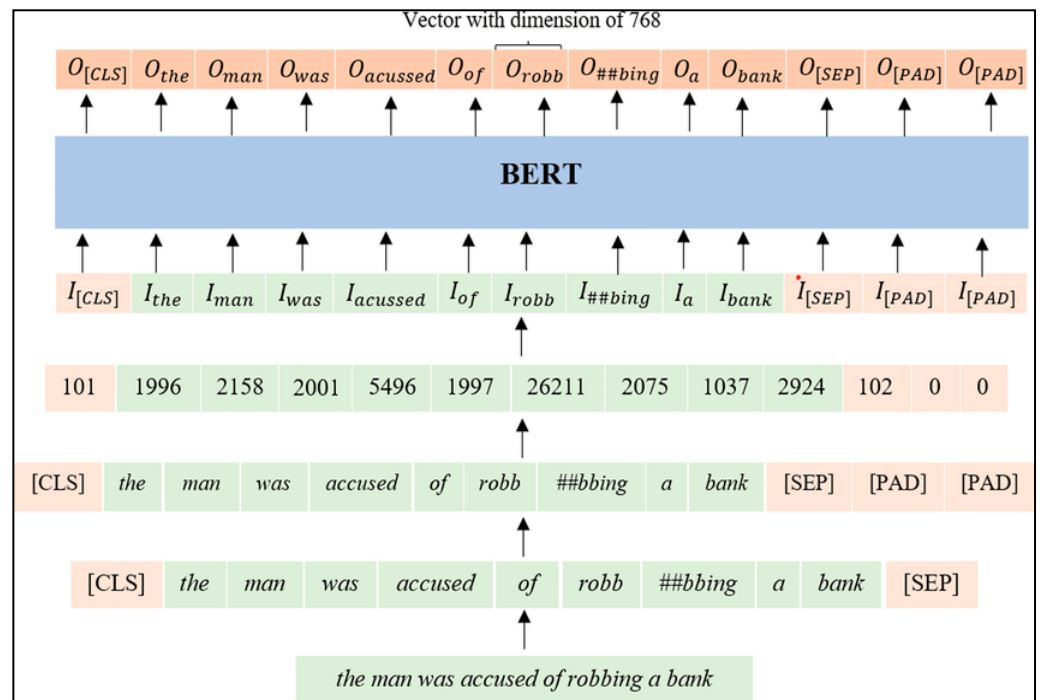


Figure 4. Illustration of text representation extraction from BERT [46].

3.4.5. Data Loading

In the experiment, the dataset was divided into two parts: the training data, which comprises 80% of the dataset, and 20% of the testing data. The training and testing dataset are loaded into the models using the data loader function where the batch size and epoch are initialized. In our case, multi-process mode is used to iterate the dataset where the epoch is 15 and the batch size is 32 (Figure 5).

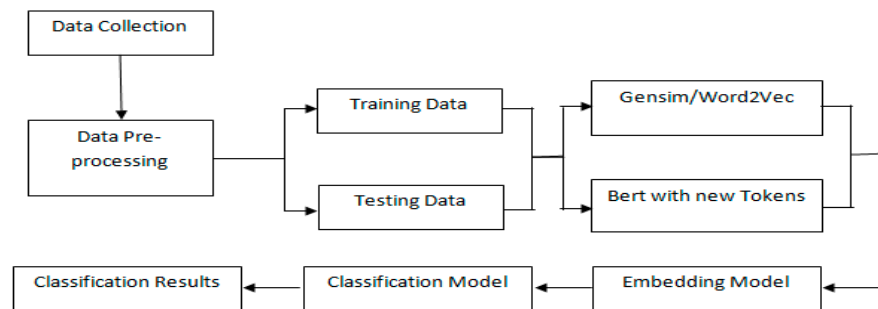


Figure 5. The overall process of the experiment.

Process: classification based on different models

- Input: the resumes of students
 - Output: A model trained on the CVs of students and one of the five pre-defined classes for each resume in the test dataset
1. Import dataset file (CVs.csv) into pandas data frame.
 2. Pre-process data (cleaning and deleting noisy data).
 3. Generate one hot encoding for each class representing the field of study.
 4. Split dataset into two parts, training and testing dataset, with ratio the 80:20, respectively.
 - a. Tokenization step based on either BERT-obtained model or the Gensim embedding approach where the tokenization was based on the unigram mode.
 5. Add new token-related competencies and unknown vocabulary into the vocab.txt of the BERT models.

6. Create an embedding matrix for every word in the vocabulary
7. Build a simple model or hybrid model based on a combination of CNN, LSTM, and GRU.
8. Dropout layer (0.2).
9. Dense (5 classes) layer with Softmax activation function.
10. Train the model on the training set.
11. Evaluate the model on the test set.

Note: The fifth step of the Gensim algorithm has been omitted.

3.4.6. Experimental Settings

The experiments were conducted in the Google Colab Pro environment, which has many computing units (100) and powerful GPUs (TPU, GPU T4, GPU A100, and GPU V100). Particularly, the BERT architecture requires high memory and GPU resources. The dataset used is the result of Extraction, Loading, and Transformation (ETL) techniques applied to raw documents, as shown in Table 3. The BERT tokenizer and associated template used are “dbmdz/bert-base-french-europeana-cased”. In particular, the majority of CVs were written in French. As a final step, we adopted the following parameters and architectures:

- Tensorflow and Keras libraries were used;
- Number of LSTM, GRU and CNN(Conv1d) layers: 1;
- Dropout rate: 0.2;
- Activation Function: SoftMax;
- $\text{learning_rate} = 1 \times 10^{-5}$;
- $\text{decay} = 1 \times 10^{-6}$;
- loss function: CategoricalCrossentropy();
- Learning rate: 0.001;
- Epochs: 15;
- Batch size: 32;
- Optimizer: Adam.

4. Results

In the first step of the experiment, we aim to verify whether the two classification methods are effective in the context of CVs, particularly for Moroccan students where such experiments are still little explored. The second step consists of comparing the two approaches, BERT and Gensim, combined with neural network approaches (GRU, LSTM, CNN) and which models are the most accurate and precise for classification.

Accordingly, based on the determined architecture adopted for this study (Figures 1 and 5), it will be possible to test and discuss the results of simple and hybrid deep learning models, which are combinations of different neural network models with two different word embedding approaches: BERT and Gensim. Accordingly, the classifications of the students' resumes were evaluated based on three parameters: accuracy, precision, and recall. However, the first observation related to model execution time; BERT models require a little more time to train and predict due to their complexity compared to Gensim models. On the other hand, according to the three evaluation measures (accuracy, precision, recall), BERT-based models are more accurate, precise, and recallable (Table 4).

As shown in Table 4, this study's results were based on three different metrics of evaluation (accuracy, precision, and recall). The first observation indicates the performance of the models based on BERT improved after including a vocabulary dictionary containing new words related to competencies and skills that were unknown in the original versions. Furthermore, compared to the Gensim-based approaches, models based on BERT text representation achieved the best accuracy and performance for the majority of approaches, especially the models based on CNN as one layer or combined with the other neural networks LSTM and GRU. Accordingly, the hybrid models based on BERT achieved the best accuracy compared to the single models from the two approaches BERT and Gensim,

where CNN—GRU/BERT has the best accuracy with 0.9351, while the GRU—LSTM/BERT has the highest precision with 0.9329. Finally, the best recall score was achieved with the CNN—GRU model based on the BERT at 0.9411.

Furthermore, the single learning models have also good metrics, especially based on BERT embedding architectures, where CNN has the best accuracy with 0.9188. On the other hand, the highest precision is achieved by the GRU with 0.9181, while LSTM/BERT achieved the best recall with 0.8963.

Table 4. Evaluation neural network approaches based on BERT/Gensim.

Text Representation Method	Model	Accuracy	Precision	Recall
BERT	GRU—LSTM	0.8122	0.8995	0.8331
	GRU—CNN	0.8821	0.8722	0.8754
	LSTM—GRU	0.8354	0.9021	0.8463
	LSTM—CNN	0.8531	0.8911	0.8234
	CNN—GRU	0.9351	0.9310	0.9411
	CNN—LSTM	0.9242	0.9329	0.9012
	GRU	0.9013	0.9181	0.8051
	LSTM	0.8951	0.8886	0.8125
	CNN	0.9188	0.9102	0.8963
	Gensim	GRU—LSTM	0.8241	0.8542
GRU—CNN		0.8321	0.8669	0.7725
LSTM—GRU		0.8214	0.8632	0.7921
LSTM—CNN		0.9025	0.8552	0.7626
CNN—GRU		0.8751	0.8224	0.7995
CNN—LSTM		0.8423	0.8821	0.7768
GRU		0.7742	0.8256	0.7951
LSTM		0.8287	0.8413	0.7858
CNN		0.9021	0.8961	0.7551

For the models based on Gensim, the three metrics also showed good performances, where the hybrid models were shown as the best models where LSTM—CNN/Gensim achieved the best accuracy with 0.9025. Meanwhile, the model CNN-LSTM/Gensim has good precision with 0.8821, whereas the model LSTM-GRU achieved good recall with 0.7995.

Finally, for the single models based on Gensim, the CNN model reached a good accuracy and precision with 0.9021 and 0.8961, respectively, while GRU had a recall value of 0.7951.

From the results obtained from the study, the hybrid models are presented based on BERT embeddings utilizing the capabilities of the GRU, LSTM, and CNN approaches. The training of the models used many layers, which is a concept of deep learning models and enhanced the models' accuracy. Accordingly, the BERT hybrid model has a better performance compared to the Gensim hybrid model as shown in the comparison in Figure 6 where the training and prediction history of models are shown. On the other hand, the model presents a good correlation between training and validation steps, while for the Gensim models, it is possible that there is more difference that can lead to the under-fitting and over-fitting issues of the models.

Using the confusion matrix as a visual tool is an essential step to have a good interpretation of the model's performance (Figure 7). A confusion matrix is a matrix that displays the number of accurate and inaccurate instances resulting from a machine learning model's predictions. This allows for a better understanding of the model's recall, accuracy, precision, and overall effectiveness in distinguishing between classes.

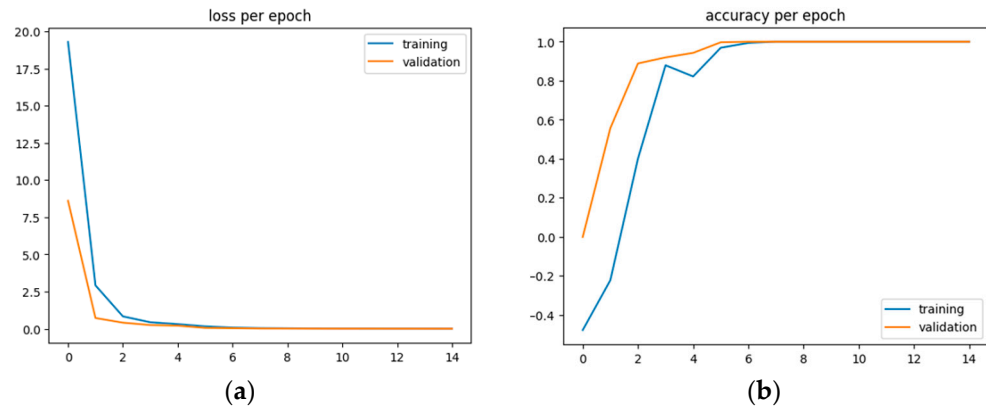


Figure 6. (a) The loss function of training and validation of resumes dataset; (b) the accuracy of training and validation of resumes dataset for the model CNN-GRU/BERT with high accuracy.

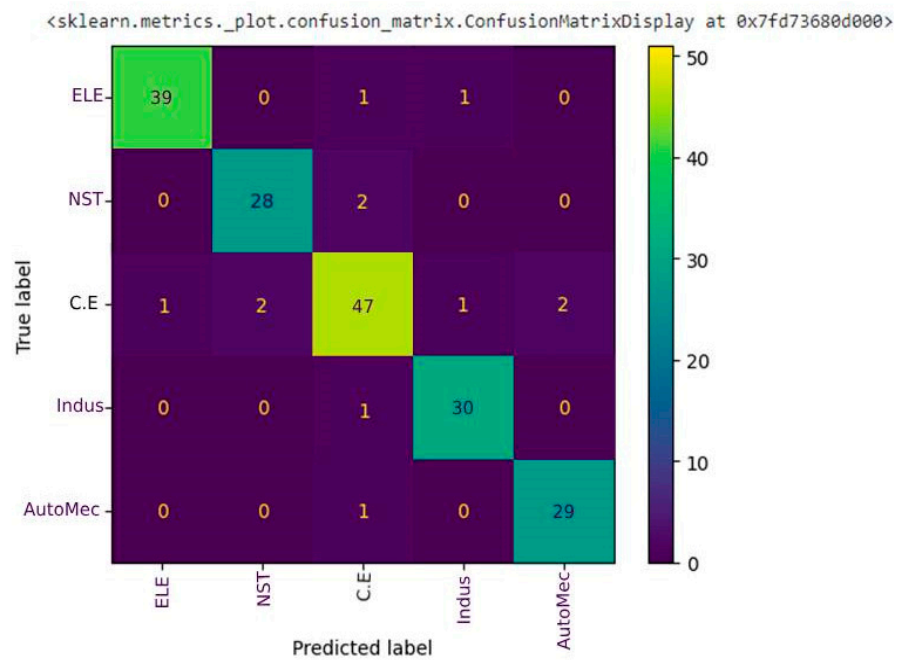


Figure 7. Confusion matrix for CV classification using the CNN-GRU/BERT model. We can explain the misclassification of the Computer Engineering with other classes due to the proximity of these categories in the uses of certain skills, in particular hard skills presented with the computer technologies.

According to the confusion matrix, CNN-GRU/BERT predicts certain classes with greater precision and recall than other models, as shown in Table 2. In accordance with the obtained results, the model predicts Electrical Engineering (ELE) with the highest score compared to other specialties with a precision of 97.5% and a recall of 95.1%; in contrast, with a precision of 90.3% and recall of 88.6%, Computer Engineering (CE) received the lowest scores among all specialties (Table 5).

Table 5. Performance outcomes related to specialty. CNN-GRU/BERT.

Specialty	Precision	Recall
Electrical Engineering (ELE)	0.975	0.951
Networks and Systems Telecommunications (NST)	0.933	0.933
Computer Engineering (CE)	0.903	0.886
Industrial Engineering (Indus)	0.937	0.967
Automotive Mechatronics Engineering (AutoMec)	0.935	0.966

The misclassifications of the Computer Engineering (CE) with other classes can be explained because the specialty of Computer Engineering has more common competencies with the other specialties, especially regarding the use of common technologies as hard skills related to IT such as “python”, “java”, and “C++”.

The actual study also showed good results compared to previous works dealing with the same subject of CV classification. Ref. [47] obtained a 78.53% accuracy of classification of resumes based on Linear SVM compared to models of comparison including Logistic Regression (62.40%), Multinomial Naïve Bayes (44.39%), and Random Forest (38.99%). In contrast, [10] based on machine learning algorithms for classification of resumes such as Naïve Bayes, Random Forest, and SVM achieved worse results presented, respectively, by Naïve Bayes (45%), SVM (60%), and Random Forest (70%).

Thus, in the Moroccan context, machine learning-based classification research is still very limited but is progressing to find its potential. Ref. [48] proposed job offer classification models for the Moroccan job market based on SVM, Naïve Bayes, Logistic Regression, and BERT. Accordingly, the BERT model has the best accuracy of 94%; however the sector classification shows a considerable difference, especially in IT-related offers, which have the lowest accuracy rate of 85% compared with different specialties' offers.

In Figure 8, we visualize the t-SNE based on the embedded method based on the text data representations from BERT. Accordingly, based on the BERT integration model, four different clusters are observed and grouped properly, unlike the Gensim model where the distinction of each group is not well observed.

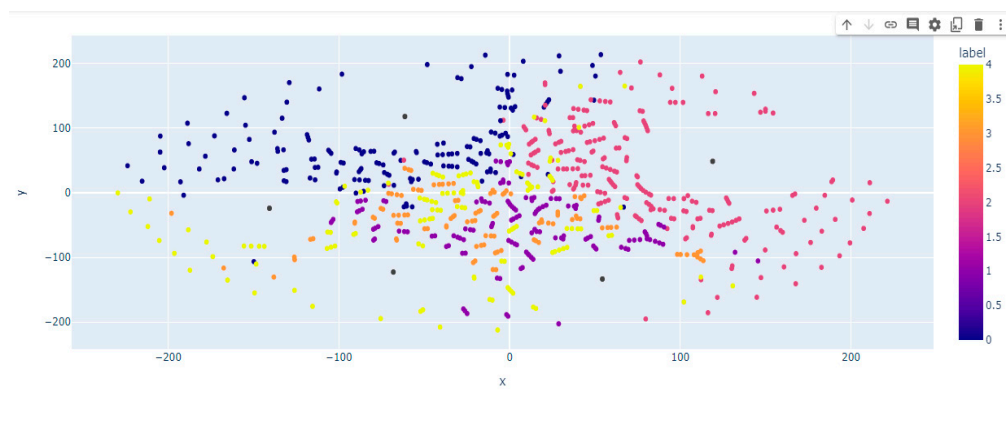


Figure 8. The t-SNE visualization of text representation of resumes using BERT and Gensim embedding methods.

5. Conclusions

The study in this paper presented the application of simple and hybrid deep learning models for the classification of CVs. Accordingly, the classifications were based on a dataset that contained 867 CVs of ENSAK students. The approach adopted for this study was based on three neural network techniques, GRU, LSTM, and CNN, in single mode or combined mode. The neural network approaches adopted were combined with two different text representation techniques generated based on BERT and Gensim text embedding techniques. We compared the performance of the hybrid models built with neural networks GRU, LSTM, and CNN using text representation from BERT with that of the embedding method Gensim respecting three metrics (accuracy, precision, and recall). We fed the embedding outputs of both algorithms (BERT and Gensim) as inputs into the neural networks in the comparative study. Therefore, CNN—GRU/BERT has the best accuracy with 0.9351 while the GRU—LSTM/BERT has the highest precision with 0.9329. Finally, the best recall score was achieved with the CNN—GRU model based on the BERT method at 0.9411. Additionally, the confusion matrix and t-SNE presentation showed good learning process of the BERT models, with an especially good interpretation for the classes

of Automotive Electrical Engineering (ELE) and Industrial Engineering (Indus), respectively, where precision reached 0.975 and 0.937. In contrast, the class Computer Engineering has misclassification with classes such as Networks and Systems Telecommunications and Electrical Engineering due to the common skills found on the related CVs, which were generally presented as IT competencies.

In conclusion, one of the reasons why BERT is effective in representing text data is that it can group related texts more closely. In general, combining deep learning methods with text representation embedding methods, such as BERT, will produce better results than single deep learning models. Finally, this study shows promising results regarding automated classification and recommendation of job applications based on CV content. With the development of real-time automated classification systems, different entities can make student recruitment processes more efficient by using relevant information such as CVs.

Author Contributions: Conceptualization, A.Q. and A.M.; methodology, A.Q.; software, A.Q.; validation, A.M. and Y.L.; formal analysis, A.Q.; investigation, A.Q. and A.M.; data curation, A.Q. and A.M.; writing—original draft preparation, A.Q.; writing—review and editing, A.Q.; visualization, A.Q.; supervision, Y.L.; project administration, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: These data are not publicly available due to ongoing research.

Conflicts of Interest: The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Nichols, J.A.; Chan, H.W.H.; Baker, M.A.B. Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophys. Rev.* **2019**, *11*, 111–118. [[CrossRef](#)] [[PubMed](#)]
- Kaul, V.; Enslin, S.; Gross, S.A. History of artificial intelligence in medicine. *Gastrointest. Endosc.* **2020**, *92*, 807–812. [[CrossRef](#)] [[PubMed](#)]
- Li, Q.; Cai, W.; Wang, X.; Zhou, Y.; Feng, D.D.; Chen, M. Medical image classification with convolutional neural network. In Proceedings of the IEEE 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), Singapore, 10–12 December 2014; pp. 844–848. [[CrossRef](#)]
- Yağcı, M. Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learn. Environ.* **2022**, *9*, 11. [[CrossRef](#)]
- Nieto, Y.; Gacia-Diaz, V.; Montenegro, C.; Gonzalez, C.C.; Crespo, R.G. Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions. *IEEE Access* **2019**, *7*, 75007–75017. [[CrossRef](#)]
- Ramteke, J.; Shah, S.; Godhia, D.; Shaikh, A. Election result prediction using Twitter sentiment analysis. In Proceedings of the IEEE 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016; pp. 1–5. [[CrossRef](#)]
- Alaei, A.R.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* **2019**, *58*, 175–191. [[CrossRef](#)]
- Golowko, N. *Future Skills in Education: Knowledge Management, AI and Sustainability as Key Factors in Competence-Oriented Education; Sustainable Management, Wertschöpfung und Effizienz*; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2021; ISBN 978-3-658-33996-8. [[CrossRef](#)]
- Huang, A.Y.Q.; Lu, O.H.T.; Huang, J.C.H.; Yin, C.J.; Yang, S.J.H. Predicting students' academic performance by using educational big data and learning analytics: Evaluation of classification methods and learning logs. *Interact. Learn. Environ.* **2020**, *28*, 206–230. [[CrossRef](#)]
- Pal, R.; Shaikh, S.; Satpute, S.; Bhagwat, S. Resume Classification using various Machine Learning Algorithms. *ITM Web Conf.* **2022**, *44*, 03011. [[CrossRef](#)]
- Urdaneta-Ponte, M.C.; Oleagordia-Ruiz, I.; Méndez-Zorrilla, A. Using LinkedIn Endorsements to Reinforce an Ontology and Machine Learning-Based Recommender System to Improve Professional Skills. *Electronics* **2022**, *11*, 1190. [[CrossRef](#)]
- Cole, M.S.; Feild, H.S.; Giles, W.F.; Harris, S.G. Recruiters' Inferences of Applicant Personality Based on Resume Screening: Do Paper People have a Personality? *J. Bus. Psychol.* **2009**, *24*, 5–18. [[CrossRef](#)]

13. Kumalasari, L.D.; Susanto, A. Recommendation System of Information Technology Jobs using Collaborative Filtering Method Based on LinkedIn Skills Endorsement. *SISFORMA* **2020**, *6*, 63–72. [[CrossRef](#)]
14. Appadoo, K.; Soonnoo, M.B.; Mungloo-Dilmohamud, Z. Job Recommendation System, Machine Learning, Regression, Classification, Natural Language Processing. In Proceedings of the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 16–18 December 2020; pp. 1–6. [[CrossRef](#)]
15. Kowsari, K.; Meimandi, J.K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text Classification Algorithms: A Survey. *Information* **2019**, *10*, 150. [[CrossRef](#)]
16. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep Learning--based Text Classification: A Comprehensive Review. *ACM Comput. Surv.* **2022**, *54*, 1–40. [[CrossRef](#)]
17. Sellamy, K.; El Farouki, M.; Sabri, Z.; Nouib, H.; Qostal, A.; Fakhri, Y.; Moumen, A. Exploring the IT's Needs in Morocco Using Online Job Ads. In *Automatic Control and Emerging Technologies*; El Fadil, H., Zhang, W., Eds.; Springer Nature: Singapore, 2024; pp. 665–677. [[CrossRef](#)]
18. Durđević Babić, I. Machine learning methods in predicting the student academic motivation. *Croat. Oper. Res. Rev.* **2017**, *8*, 443–461. [[CrossRef](#)]
19. Qazdar, A.; Er-Raha, B.; Cherkaoui, C.; Mammass, D. A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco. *Educ. Inf. Technol.* **2019**, *24*, 3577–3589. [[CrossRef](#)]
20. Mourdi, Y.; Sadgal, M.; Berrada Fathi, W.; El Kabtane, H. A Machine Learning Based Approach to Enhance Mooc Users' Classification. *Turk. Online J. Distance Educ.* **2020**, *21*, 47–68. [[CrossRef](#)]
21. Sadqui, A.; Ertel, M.; Sadiki, H.; Amali, S. Evaluating Machine Learning Models for Predicting Graduation Timelines in Moroccan Universities. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 10. [[CrossRef](#)]
22. Ouatik, F.O.; Erritali, M.E.; Jourhmane, M.J. Student orientation using machine learning under MapReduce with Hadoop. *J. Ubiquitous Syst. Pervasive Netw.* **2020**, *13*, 21–26. [[CrossRef](#)]
23. Qostal, A.; Moumen, A.; Lakhrissi, Y. Systematic Literature Review on Big Data and Data Analytics for Employment of Youth People: Challenges and Opportunities. In *Proceedings of the 2nd International Conference on Advanced Technologies for Humanity*; SCITEPRESS—Science and Technology Publications: Rabat, Morocco, 2020; pp. 179–185. [[CrossRef](#)]
24. Casuat, C.D.; Festijo, E.D. Predicting Students' Employability using Machine Learning Approach. In Proceedings of the 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia, 20–21 December 2019; pp. 1–5. [[CrossRef](#)]
25. Mewburn, I.; Grant, W.J.; Suominen, H.; Kizimchuk, S. A Machine Learning Analysis of the Non-academic Employment Opportunities for Ph.D. Graduates in Australia. *High. Educ. Policy* **2020**, *33*, 799–813. [[CrossRef](#)]
26. ElSharkawy, G.; Helmy, Y.; Yehia, E. Employability Prediction of Information Technology Graduates using Machine Learning Algorithms. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*. [[CrossRef](#)]
27. Roy, A. Recent Trends in Named Entity Recognition (NER). *arXiv* **2021**. [[CrossRef](#)]
28. Narendra, G.O.; Hashwanth, S. Named Entity Recognition based Resume Parser and Summarizer. *Int. J. Adv. Res. Sci. Commun. Technol.* **2022**, *2*, 728–735. [[CrossRef](#)]
29. Gughani, A.; Misra, H. Implicit Skills Extraction Using Document Embedding and Its Use in Job Recommendation. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 13286–13293. [[CrossRef](#)]
30. Fareri, S.; Melluso, N.; Chiarello, F.; Fantoni, G. SkillNER: Mining and mapping soft skills from any text. *Expert Syst. Appl.* **2021**, *184*, 115544. [[CrossRef](#)]
31. Casuat, C.D. Predicting Students' Employability using Support Vector Machine: A SMOTE-Optimized Machine Learning System. *Int. J. Emerg. Trends Eng. Res.* **2020**, *8*, 2101–2106. [[CrossRef](#)]
32. Baffa, M.H.; Miyim, M.A.; Dauda, A.S. Machine Learning for Predicting Students' Employability. *UMYU Sci.* **2023**, *2*, 001–009. [[CrossRef](#)] [[PubMed](#)]
33. Sun, T.; He, Z. Developing intelligent hybrid DNN model for predicting students' employability—A Machine Learning approach. *J. Educ. Humanit. Soc. Sci.* **2023**, *18*, 235–248. [[CrossRef](#)]
34. Makdoun, I.; Mezzour, G.; Carley, K.M.; Kassou, I. Analyzing the Needs of the Automotive Job Market in Morocco. In Proceedings of the 2018 13th International Conference on Computer Science & Education (ICCSE), Colombo, Sri Lanka, 8–11 August 2018; pp. 1–6.
35. Habous, A.; Nfaoui, E.H. Combining Word Embeddings and Deep Neural Networks for Job Offers and Resumes Classification in IT Recruitment Domain. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 7. [[CrossRef](#)]
36. Mgarbi, H.; Chkouri, M.; Tahiri, A. Towards a New Job Offers Recommendation System Based on the Candidate Resume. *Int. J. Comput. Digit. Syst.* **2023**, *14*, 31–38. [[CrossRef](#)] [[PubMed](#)]
37. Qostal, A.; Sellamy, K.; Sabri, Z.; Nouib, H.; Lakhrissi, Y.; Moumen, A. Perceived employability of moroccan engineering students: A PLS-SEM approach. *Int. J. Instr.* **2024**, *17*, 259–282. [[CrossRef](#)]
38. Hopfield, J.J. Brain, neural networks, and computation. *Rev. Mod. Phys.* **1999**, *71*, S431–S437. [[CrossRef](#)]
39. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
40. Li, W.; Wu, H.; Zhu, N.; Jiang, Y.; Tan, J.; Guo, Y. Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Inf. Process. Agric.* **2021**, *8*, 185–193. [[CrossRef](#)]

41. Ren, L.; Cheng, X.; Wang, X.; Cui, J.; Zhang, L. Multi-scale Dense Gate Recurrent Unit Networks for bearing remaining useful life prediction. *Future Gener. Comput. Syst.* **2019**, *94*, 601–609. [[CrossRef](#)]
42. Nosouhian, S.; Nosouhian, F.; Khoshouei, A.K. A Review of Recurrent Neural Network Architecture for Sequence Learning: Comparison between LSTM and GRU. *Preprints* **2021**, 2021070252. [[CrossRef](#)]
43. O’Shea, K.; Nash, R. An Introduction to Convolutional Neural Networks. *arXiv* **2015**. [[CrossRef](#)]
44. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [[CrossRef](#)]
45. Alaparthi, S.; Mishra, M. Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. *arXiv* **2020**. [[CrossRef](#)]
46. Subakti, A.; Murfi, H.; Hariadi, N. The performance of BERT as data representation of text clustering. *J. Big Data* **2022**, *9*, 15. [[CrossRef](#)] [[PubMed](#)]
47. Roy, P.K.; Chowdhary, S.S.; Bhatia, R. A Machine Learning approach for automation of Resume Recommendation system. *Procedia Comput. Sci.* **2020**, *167*, 2318–2327. [[CrossRef](#)]
48. Rahhal, I.; Carley, K.M.; Kassou, I.; Ghogho, M. Two Stage Job Title Identification System for Online Job Advertisements. *IEEE Access* **2023**, *11*, 19073–19092. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.