

Article

Not peer-reviewed version

---

# Resumes Classification Using Neural Network Approaches Combined with Bert and Gensim: CVS of Moroccan Engineering Students

---

[Aniss Qostal](#)\*, [Aniss Moumen](#)\*, [Younes Lakhrissi](#)\*

Posted Date: 4 March 2024

doi: 10.20944/preprints202403.0111.v1

Keywords: Gated Recurrent Unit (GRU); Long Short-Term Memory (LSTM); Convolutional Neural Networks (CNN); BERT; Gensim; Moroccan engineering students; Ibn Tofail University; Resumes; CVS; ENSAK



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

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

Aniss Qostal <sup>1,\*</sup>, Aniss Moumen <sup>2</sup> and Younes Lakhri <sup>1</sup>

<sup>1</sup> Intelligent Systems, Georesources and Renewable Energies Laboratory (SIGER IN FRENCH), Sidi Mohamed Ben Abdellah University, FST Fez, Morocco

<sup>2</sup> Laboratory of Engineering Sciences, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco

\* 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 corpora 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 (CNN). 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 resumes 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, Networks and Systems Telecommunications, Computer Engineering, Automotive Mechatronics Engineering, Industrial Engineering). The results revealed good performance of the proposed models based on the BERT embedding approach compared to models based on the Gensim Word2Vec embedding approach. Accordingly, the CNN-GRU/Bert model achieved slightly better accuracy with 0.9251 compared to other hybrid models.

**Keywords:** Gated Recurrent Unit (GRU); Long Short-Term Memory (LSTM); Convolutional Neural Networks (CNN); BERT; Gensim; Moroccan engineering students; Ibn Tofail University; resumes; CVs; ENSAK

## 1. Introduction

The use of AI (Artificial Intelligence) including Deep Learning (DL) and Machine Learning (ML) approaches to process large corpora, 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), 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 Neural Network (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 resumes 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' resumes, 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.

## 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 [5,17].

Several studies have been carried out in the Moroccan educational context, where [18] 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 [18]. Furthermore, the author [19] 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%) [19].

In addition, the author [20] 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 [21] 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 [22] 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 [23,24]. To build their models, different factors are taken into account, including technical skills, soft skills, personality traits, demographics, extracurricular activities, and internships[25]. Hence, different sources are considered, including recruitment platforms, professional social media platforms, CV platforms, and unstructured CV files [25]. 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.) [25].

The first approach involves extracting relevant information from different corpora and detecting features by using Named Entity Recognition (NER) and Named Entity Normalization (NEN)[26]. 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 [27]. 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[28,29].

**Table 1.** Predicting employability models using different approaches of Machine learning.

Study	Context	Model	accuracy
[23]	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 Trees (DT)	- DT :85%
		- Random Forest (RF)	-RF :84%
		- Support vector machine (SVM)	-SVM:92.22%
[30]	Dataset based on Mock job Interview Results with three thousand (3000) observations and twelve (12) features, Student Performance Rating of the OJT students collected	- Decision Trees (DT)	- DT : 56.36%
		- Random Forest (RF)	- RF: 64%
		- Support vector machine (SVM)	-SVM :91.22 %
		-K- Nearest Neighbor (KNN)	-KNN:79.99%
		- Logistic Regression (LR)	-LR: 65.45%

[25]	Dataset (296 records)	- Decision tree (DT)	- DT : 100%
	from survey IT graduates	- Gaussian NB	- NB : 92%
	and employers in Egypt	-Logistic Regression (LR)	- LR : 98%
	-Training skills	- Random Forest (RF)	- RF: 97%
	-soft skills	-Support Vector Machine	- SVM : 98%
	-hard skills	(SVM)	
[31]	Proposed models for	- Random Forest RF	-RF : 98%
	predict performance and	- Logistic Regression LR	-LR: 94%
	students' employability.	- Decision Tree	-DT: 97%
	Primary datasets of 218		
	graduate students of		
	higher educational		
	institutions (HEIs).		
[32]	Hybrid DNN model for	- Random Forest (RF)	-SVM : 92%
	predicting students'	-Support Vector Machine	-RF: 89%
	employability – A Machine	(SVM)	-GB: 84%
	Learning approach		-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 [33] 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, [34] 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 [35] 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 resumes dataset was collected as part of a survey on the subject of employability of the Moroccan students, in our case, engineering students from ENSA Kenitra (Table 2). Additionally, the study was conducted as a continuation of previous research related to students' skills and their perceptions of labor market demands [36]. Therefore, the ENSAs 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.

The dataset contained 867 resumes from five departments (Electrical Engineering, Networks and Systems Telecommunications, Computer Engineering, Automotive Mechatronics Engineering, Industrial Engineering). 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.

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

ENSA Kenitra	Department	Total
Department	Computer Engineering	263
	Networks and Systems Telecommunications	134
	Automotive Mechatronics Engineering	149
	Industrial Engineering	114
	Electrical Engineering	207
Total		867

### 3.2. Experiment and Problem Definition

In the first step, the experiment started by loading the collected dataset which is 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 are 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 was 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 is loaded as input for the different neural networks adopted in the research including GRU, LSTM, and CNN.

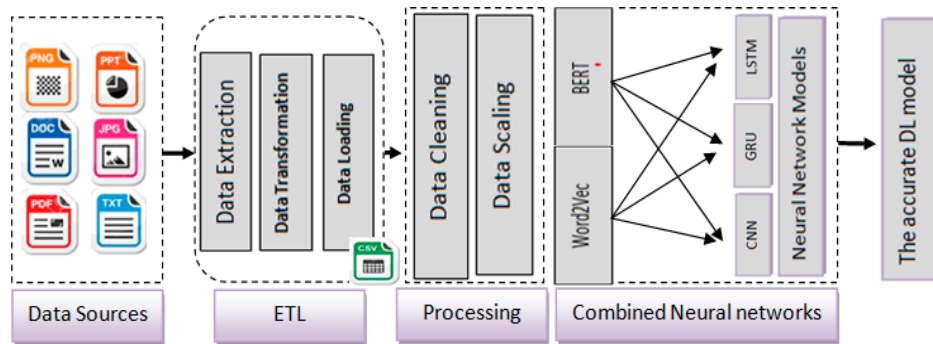
**Table 3.** File types that make up the dataset of student resumes.

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  $C = \{c1, ..., cn\}$  and the student's specialty as  $S = \{s1, ..., sk\}$  as the targets. The models classified and measured the probabilities  $\{(c1, s1), ..., (cn, sk)\}$ . The input to the classifier is the training data, which is a finite sequence of  $S \times C$  pairs. The output of the classifiers 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 [37]. 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 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], [PAD]. The text classification system can be decomposed into the following four stages: Feature extraction, dimension reductions, classifier selection, and evaluation.

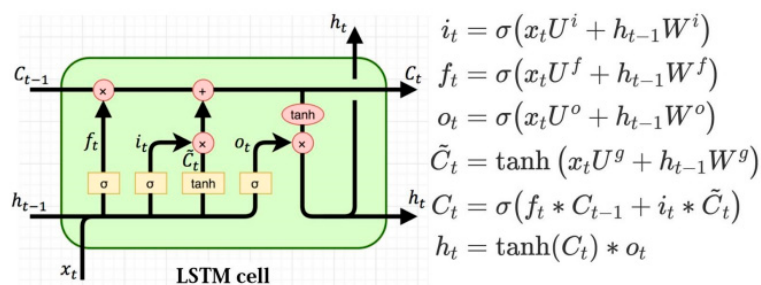


**Figure 1.** Combined Neural network models with Bert/Gensim.

### 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 [38]. 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 nonlinear activation functions such as Tanh or Sigmoid. This solution enables LSTM models to store and transmit information over the long term. However, the standard formula for a single LSTM unit can be given by the following equation [37,38]:



**Figure 2.** Cell structure and equations describing LSTM gates.

Here  $\sigma$  is the sigmoid function,  $\tanh$  is the hyperbolic tangent function,  $i$ ,  $f$ ,  $o$ ,  $C$ ,  $C_{\sim}$  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 Neural Network (GRU)

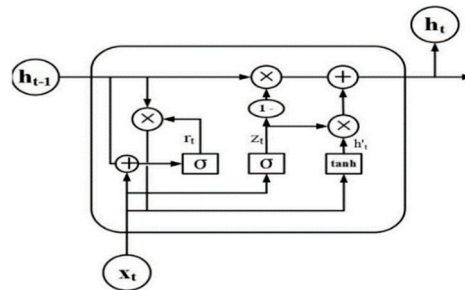
Gated Recurrent Unit (GRU) was introduced in 2014 as a solution to LSTM's complexity and as a solution to the vanishing gradient problem [41,42]. Moreover, by implementing gating mechanisms within their networks, GRU and LSTM can capture and propagate information over long sequences. Thus, GRU consists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

**Update Gate (zt):** Calculates how much information from the past should be carried forward.

**Reset Gate (rt):** It determines how much information should be forgotten about the past.

**Current Memory Content (ht):** 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.



**Figure 3.** Diagram of a one-unit Gated Recurrent Unit (GRU).

### 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 consist 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, pooling, and convolutional 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 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova first published "All Attention You Need" in 2018 at Google Labs [42]. The model is an embedding layer of pre-trained bidirectional representations from a large collection of unsupervised text corpora, including Wikipedia and BookCorpus [42]. 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:

#### 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 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; "[UNK]" to indicate the sequence in the sequence unknown word

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 [42]. 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.

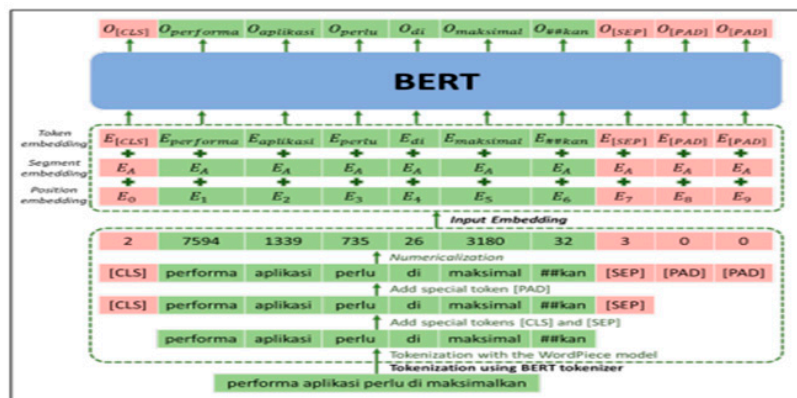


Figure 4. Bert Architecture .

#### 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 is 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.

#### Algorithm: classification based on different models

- Input: The Resume of students
- Output: A model trained on the Resume students and one of the 5 pre-defined classes for each resume in the test dataset

1. Import dataset file (Resumes.csv) into pandas dataframe.
2. Pre-processing data (cleaning and deleting noisy data).
3. Generate one hot encoding for each class representing the filed study.
4. Split dataset into two parts training and testing dataset with ratio 80:20 respectively.
  - a. Tokenization step based on either Bert pertained model or the Gensim embedding approach.
5. Adding 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. Building 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 more 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
- learning\_rate=1e-5
- decay=1e-6
- loss function: CategoricalCrossentropy()
- Learning rate: 0.001
- Epochs: 15
- Batch size: 32
- Optimizer: Adam

## 4. Results

Based on the figured architecture adopted for the study (Figure 1 and Figure 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).

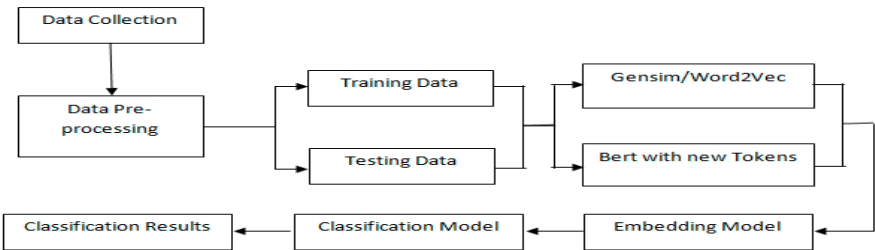


Figure 5. : the overall process of the experiment.

Table 4. Evaluation Neural Network Approaches Based on Bert/Gensim.

Text Representation Method	Model	Accuracy	Precision	Recall
BERT	GRU - LSTM	0.7522	0.9050	0.8331
	GRU - CNN	0.8821	0.8722	<b>0.8481</b>
	LSTM - GRU	0.7214	0.9021	0.8123
	LSTM - CNN	0.7334	0.8911	0.8234
	CNN - GRU	<b>0.9251</b>	0.8541	0.8442
	CNN - LSTM	0.8423	<b>0.9084</b>	0.8314
	GRU	0.9013	<b>0.9181</b>	0.8051
	LSTM	0.8951	0.8886	<b>0.8125</b>
	CNN	<b>0.9141</b>	0.9015	0.8121
	GRU - LSTM	0.8241	0.8542	0.7741
	GRU -CNN	0.8321	0.8669	0.7725
Gensim	LSTM - GRU	0.8214	0.8632	0.7921
	LSTM - CNN	<b>0.9025</b>	0.8552	0.7626
	CNN - GRU	0.8751	0.8224	<b>0.7995</b>
	CNN - LSTM	0.8423	<b>0.8821</b>	0.7768
	GRU	0.7742	0.8256	<b>0.7951</b>
	LSTM	0.8287	0.8413	0.7858
	CNN	<b>0.9021</b>	<b>0.8961</b>	0.7551

As shown in Table 4, the 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 0.9251, while the GRU – LSTM/BERT has the highest precision with 0.9084. Finally, the best recall score was achieved with the GRU-CNN model based on the BERT 0.8481.

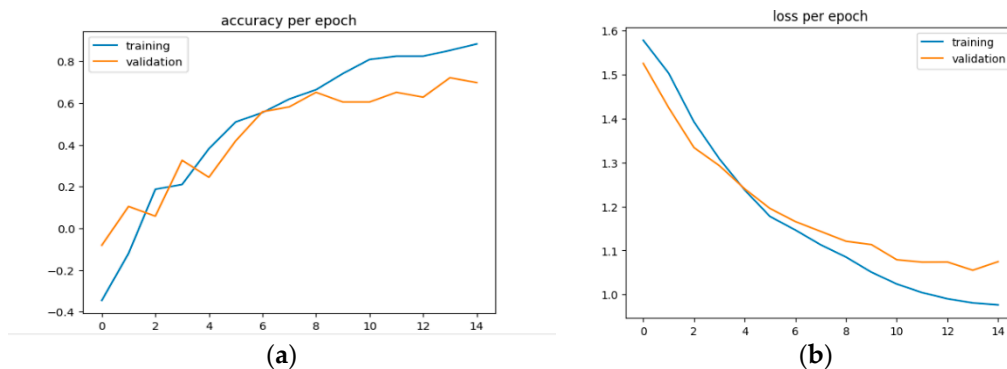
Furthermore, the single learning models have also good metrics especially based on Bert embedding architectures, where CNN has the best accuracy with 0.9141. On the other hand, the highest precision is achieved by the GRU with 0.9015, while LSTM/Bert achieved the best recall with 0.8892.

For the models based on Gensim, the three metrics also showed good performances, where the hybrid models were shown as well as the best models where LSTM – CNN/Gensim achieved the best

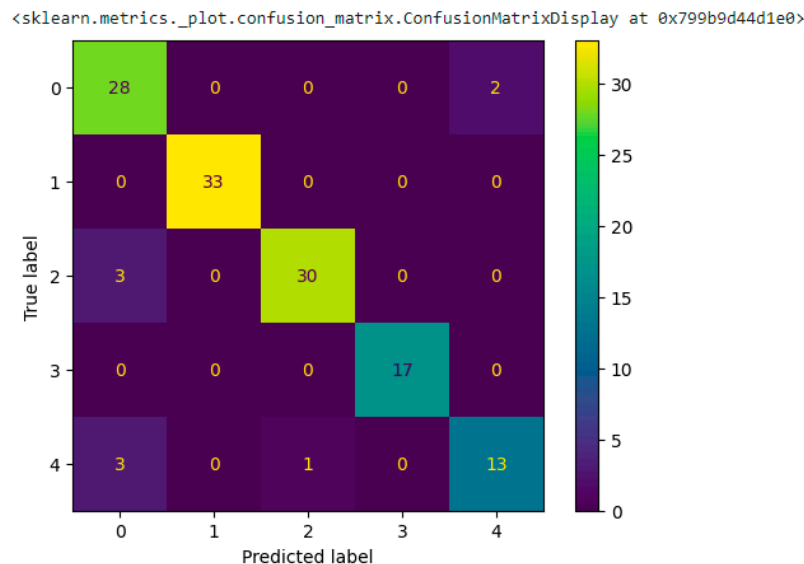
accuracy with 0.9025. Meanwhile, the model CNN-LSTM /Gensim has good precision with 0.8821, whereas the model LSTM-GRU achieved good recall with 0.7995.

Finally, the single models based on Gensim, the CNN model reached 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 presented based on BERT embeddings utilizing the capabilities of GRU, LSTM, and CNN neural networks approaches where the training of the models using many layers is a concept of deep learning model and enhancer of 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 presented a good correlation between training and validation steps, while for the gensim models, it is constable there is more difference which can lead to the under-fitting and over-fitting issues for the models.



**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.

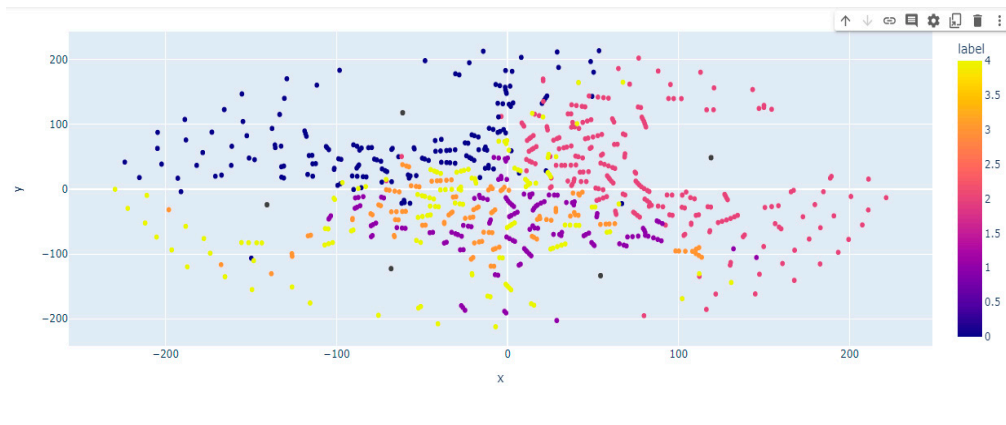


**Figure 7.** Confusion matrix model CNN-GRU/Bert.

Using the confusion matrix as a visual tool is an essential step to have a good interpretation of the model's performance. Therefore, a confusion matrix is a matrix that displays the number of accurate and inaccurate instances resulting from a machine learning model's predictions. Accordingly, this analysis provides a deep examination of the predictions made by a model, including True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). This allows for a better understanding of the model's recall, accuracy, precision, and overall effectiveness in distinguishing between classes.

The obtained results based on the confusion matrix, where the presented model CNN-GRU/Bert, indicated good interpretation for the classe 1 and 3 representing Industriel and automotive respectively, otherwise the classe 0 has confusion with classe 2 , 4 representing Computer Engineering, Networks and Systems Telecommunications and Electrical Engineering respectively. Where the speciality of computer science has more commun competencies with the other specialities, especially on the use of commun technologies on the education process.

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 the paper presented the application of simple and hybrid deep learning models. Accordingly, the classification of the resumes of ENSAK students was based on text representation 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 0.9251 while the GRU – LSTM/BERT has the highest precision with 0.9084. Finally, the best recall score was achieved with the GRU-CNN model based on the BERT 0.8481. Additionally, the confusion matrix and t-sne presentation showed good learning process of the Bert models especially good interpretation for the classe 1 and 3 representing Automotive Mechatronics Engineering and Industrial Engineering respectively, otherwise the classe 0 has confusion with classes 2 and 4 and 0 representing Computer Engineering , Networks and Systems Telecommunications and Electrical Engineering respectively

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, it is evident that more research on employability is needed, using machine learning to uncover students' personality traits, skill extraction, and matching skills to required professional qualities.

## References

1. Nichols, J.A.; Herbert Chan, H.W.; Baker, M.A.B. Machine Learning: Applications of Artificial Intelligence to Imaging and Diagnosis. *Biophys. Rev.* 2019, 11, 111–118, doi:10.1007/s12551-018-0449-9.



2. Kaul, V.; Enslin, S.; Gross, S.A. History of Artificial Intelligence in Medicine. *Gastrointest. Endosc.* 2020, 92, 807–812, doi:10.1016/j.gie.2020.06.040.
3. Li, Q.; Cai, W.; Wang, X.; Zhou, Y.; Feng, D.D.; Chen, M. Medical Image Classification with Convolutional Neural Network. In Proceedings of the 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV); IEEE: Singapore, December 2014; pp. 844–848.
4. Yağcı, M. Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms. *Smart Learn. Environ.* 2022, 9, 11, doi:10.1186/s40561-022-00192-z.
5. Nieto, Y.; Gacia-Diaz, V.; Montenegro, C.; Gonzalez, C.C.; Gonzalez Crespo, R. Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions. *IEEE Access* 2019, 7, 75007–75017, doi:10.1109/ACCESS.2019.2919343.
6. Ramteke, J.; Shah, S.; Godhia, D.; Shaikh, A. Election Result Prediction Using Twitter Sentiment Analysis. In Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT); IEEE: Coimbatore, India, August 2016; pp. 1–5.
7. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* 2019, 58, 175–191, doi:10.1177/0047287517747753.
8. 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, 2021; ISBN 978-3-658-33996-8.
9. Mao, Y.; Chi, M.; Lin, C. Deep Learning vs. Bayesian Knowledge Tracing: Student Models for Interventions. 2018, 10.
10. Pal, R.; Shaikh, S.; Satpute, S.; Bhagwat, S. Resume Classification Using Various Machine Learning Algorithms. *ITM Web Conf.* 2022, 44, 03011, doi:10.1051/itmconf/20224403011.
11. 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, doi:10.3390/electronics11081190.
12. 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, doi:10.1007/s10869-008-9086-9.
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, doi:10.24167/sisforma.v6i2.2240.
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); IEEE: Gold Coast, Australia, December 16 2020; pp. 1–6.
15. Kowsari; Jafari Meimandi; Heidarysafa; Mendu; Barnes; Brown Text Classification Algorithms: A Survey. *Information* 2019, 10, 150, doi:10.3390/info10040150.
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, doi:10.1145/3439726.
17. Faculty of Education, University of Osijek, Cara Hadrijana 10, 31 000 Osijek, Croatia; Đurđević Babić, I. Machine Learning Methods in Predicting the Student Academic Motivation. *Croat. Oper. Res. Rev.* 2017, 8, 443–461, doi:10.17535/corr.2017.0028.
18. 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, doi:10.1007/s10639-019-09946-8.
19. 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, 47–68, doi:10.17718/tojde.727976.
20. 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, doi:10.14569/IJACSA.2023.0140734.
21. 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, doi:10.5383/JUSPN.13.01.003.
22. Qostal, A.; Moumen, A.; Lakhriissi, Y. Systematic Literature Review on Big Data and Data Analytics for Employment of Youth People: Challenges and Opportunities: In Proceedings of the Proceedings of the 2nd

- International Conference on Advanced Technologies for Humanity; SCITEPRESS - Science and Technology Publications: Rabat, Morocco, 2020; pp. 179–185.
23. Casuat, C.D. Predicting Students' Employability Using Machine Learning Approach.
  24. 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, doi:10.1057/s41307-018-0098-4.
  25. ElSharkawy, G.; Helmy, Y.; Yehia, E. Employability Prediction of Information Technology Graduates Using Machine Learning Algorithms. *Int. J. Adv. Comput. Sci. Appl.* 2022, 13, doi:10.14569/IJACSA.2022.0131043.
  26. Roy, A. Recent Trends in Named Entity Recognition (NER) 2021.
  27. Narendra G O; Hashwanth S Named Entity Recognition Based Resume Parser and Summarizer. *Int. J. Adv. Res. Sci. Commun. Technol.* 2022, 728–735, doi:10.48175/IJAR SCT-3029.
  28. Guagnani, A.; Misra, H. Implicit Skills Extraction Using Document Embedding and Its Use in Job Recommendation. *Proc. AAAI Conf. Artif. Intell.* 2020, 34, 13286–13293, doi:10.1609/aaai.v34i08.7038.
  29. Fareri, S.; Melluso, N.; Chiarello, F.; Fantoni, G. SkillNER: Mining and Mapping Soft Skills from Any Text. *Expert Syst. Appl.* 2021, 184, 115544, doi:10.1016/j.eswa.2021.115544.
  30. 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, doi:10.30534/ijeter/2020/102852020.
  31. Muhammad Hadiza Baffa; Muhammad Abubakar Miyim; Abdullahi Sani Dauda Machine Learning for Predicting Students' Employability. *UMYU Sci.* 2023, 2, 001–009, doi:10.56919/usc.2123\_001.
  32. 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, doi:10.54097/ehss.v18i.11000.
  33. 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); IEEE: Colombo, August 2018; pp. 1–6.
  34. 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, doi:10.14569/IJACSA.2021.0120774.
  35. 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, doi:10.12785/ijcds/140103.
  36. Qostal, A.; Sellamy, K.; Sabri, Z.; Nouib, H.; Lakhri, Y.; Moumen, A. Perceived Employability of Moroccan Engineering Students: A PLS-SEM Approach. *Int. J. Instr.* 2024, 17, 259–282.
  37. Hopfield, J.J. Brain, Neural Networks, and Computation. *Rev. Mod. Phys.* 1999, 71, S431–S437, doi:10.1103/RevModPhys.71.S431.
  38. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* 1997, 9, 1735–1780.
  39. Rana, M.; Uddin, Md.M.; Hoque, Md.M. Effects of Activation Functions and Optimizers on Stock Price Prediction Using LSTM Recurrent Networks. In Proceedings of the Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence; ACM: Normal IL USA, December 6 2019; pp. 354–358.
  40. Varsamopoulos, S.; Bertels, K.; Almudever, C. *Designing Neural Network Based Decoders for Surface Codes*; 2018;
  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, doi:10.1016/j.future.2018.12.009.
  42. Nosouhian, S.; Nosouhian, F.; Kazemi Khoshouei, A. *A Review of Recurrent Neural Network Architecture for Sequence Learning: Comparison between LSTM and GRU*; other, 2021;
  43. O'Shea, K.; Nash, R. *An Introduction to Convolutional Neural Networks* 2015.
  44. Alaparthi, S.; Mishra, M. Bidirectional Encoder Representations from Transformers (BERT): A Sentiment Analysis Odyssey. *ArXiv Prepr. ArXiv200701127* 2020.

**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.