Paper: **"A Comprehensive Review of the Three Main Topic Modeling Algorithms and Challenges in Albanian Employability Skills"**

Corresponding Author: Milena Shehu

Peer review:

Reviewer 1: Gibet Tani Hicham
Abdemalek Essaadi University, Morocco

Reviewer 2: Alaa Alshamasneh
Prince Sultan University Kingdome of Saudi Arabia

Reviewer 3: Xi Chen
University of Kentucky, USA

Reviewer A:
Recommendation: Accept Submission

The TITLE is clear and it is adequate to the content of the article.

Yes, based on the title "A Comprehensive Review of the Three Main Topic Modeling Algorithms and Challenges in Albanian Employability Skills," it appears to be clear and adequately reflective of the content of the article. The title effectively communicates the main focus of the research, which is to provide a detailed examination of topic modeling algorithms and their challenges specifically in the context of Albanian employability skills.

The title indicates that the paper will cover three main topic modeling algorithms, discuss their application to employability skills, and address the challenges associated with this analysis in the Albanian context. This alignment between the title and the content of the article is essential for setting accurate expectations for readers and ensuring that the research objectives are clearly conveyed.

Overall, the title seems appropriate for the content of the article, as it succinctly captures the key aspects of the study and provides a clear indication of the scope and focus of the research.

The ABSTRACT clearly presents objects, methods, and results.

Objectives: The abstract clearly outlines the objectives of the study, which include analyzing the challenges faced by contemporary jobseekers in aligning their interests, employability soft skills, and professional experience with available career opportunities. The focus on the use of topic modeling algorithms in analyzing job advertisements and employability soft skills in the Albanian job market is also highlighted as a key objective.

Methods: The abstract briefly mentions the methodology employed in the study, specifically focusing on the analysis of job advertisements posted online using unsupervised learning techniques, particularly Topic Modeling algorithms such as Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), and BERTopic. The mention of comparing these methodologies indicates a comparative analysis approach.

Results: While the abstract does not provide specific numerical results, it does mention that the study analyzes and compares the three primary methodologies and algorithms used in topic modeling for employability soft skills. It also suggests that the study draws conclusions on the superior performance, optimal algorithm applicability, as well as challenges and limitations based on the review of studies conducted in the Albanian job market.

Overall, the abstract effectively presents the objectives, methods, and a general overview of the results of the study. It provides a clear understanding of the research focus, the methodology used, and the intended outcomes of the analysis of topic modeling algorithms in the context of employability soft skills in Albania.

There are a few grammatical errors and spelling mistakes in this article.

Based on the provided text, there are no mentions of grammatical errors or spelling mistakes in the article. The text mainly focuses on the challenges of using certain topic modeling algorithms in the Albanian language and the initiatives to analyze employability skills in Albania.

The study METHODS are explained clearly.

Yes, the study methods are explained clearly in the text. The text provides detailed explanations of the three main topic modeling algorithms used in the study: LSA (Latent Semantic Analysis), LDA (Latent Dirichlet Allocation), and BERTopic. It discusses the strengths, applications, and challenges of each method in analyzing soft skills in job vacancies. Additionally, it describes the preprocessing steps involved in data preparation for the analysis, such as normalization, tokenization, removal of punctuation and special characters, elimination of stop words, stemming, and corpus construction (source: '17939-article-text-51796-1-4-20240327_1712098600.pdf', page 7).

The body of the paper is clear and does not contain errors.

Based on the provided text, the body of the paper is clear and does not contain errors.

The CONCLUSION or summary is accurate and supported by the content.

Based on the provided text, the conclusion or summary is accurate and supported by the content. The text discusses the comparison of Topic Modelling Algorithms used to analyze employability soft skills, emphasizing the challenges and benefits of each algorithm. It also highlights the importance of these methodologies in the context of the Albanian job market and the potential benefits they offer in understanding employability trends and supporting decision-making in workforce development.

The list of REFERENCES is comprehensive and appropriate.

The list of references provided in the text is comprehensive and appropriate, covering various studies, articles, and sources related to topics such as youth employment, skills development, and machine learning applications in the job market.

Please rate the TITLE of this paper.

[Poor] **1-5** [Excellent]

3

Please rate the ABSTRACT of this paper.

[Poor] **1-5** [Excellent]

3

Please rate the LANGUAGE of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the METHODS of this paper.

[Poor] **1-5** [Excellent]

3

Please rate the BODY of this paper.

[Poor] **1-5** [Excellent]

3

Please rate the CONCLUSION of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the REFERENCES of this paper.

[Poor] **1-5** [Excellent]

3

Overall Recommendation!!!

Accepted, minor revision needed

Comments and Suggestions to the Author(s):

Language Complexity: The complexity of the Albanian language, including morphology, syntax, and semantics, may pose challenges for models trained on languages with different linguistic structures. Provide insights into how you accounted for these linguistic complexities in your analysis and interpretation of results.

Model Adaptation: Adapting pre-trained models like BERT to the Albanian language may require extensive fine-tuning on Albanian-specific data. Discuss the process of fine-tuning these models and any potential implications for the accuracy and reliability of your results.

Domain Specificity: Employability skills and job vacancies in Albania may have specific linguistic nuances and terminologies. Explain how you addressed these nuances in your analysis to ensure the relevance and applicability of your findings to the Albanian job market.

Evaluation and Validation: Developing custom evaluation metrics and validation procedures tailored to the characteristics of Albanian text data is crucial. Provide details on how you validated the performance of your models and the reliability of your results in the context of the Albanian language.

Methodological Transparency: Ensure that you provide a clear and detailed description of your methodology, including data preprocessing steps, model selection criteria, and validation procedures. Transparency in methodology enhances the reproducibility and credibility of your research.

Discussion of Results: In your discussion section, critically evaluate the implications of your findings in the context of the Albanian job market. Address the practical significance of your results and how they can inform policy-making or educational initiatives related to youth employability.

By addressing these limitations and incorporating feedback into your paper, you can strengthen the quality and impact of your research.

Reviewer B:
Recommendation: Accept Submission

The TITLE is clear and it is adequate to the content of the article.

The title is clear. The literature review is comprehensive but could be better organized. It might be helpful to categorize the studies reviewed into themes or methodologies to provide a clearer overview of the field.
Some of the studies mentioned in the literature review lack a direct connection to the main focus of the paper, which is the analysis of employability soft skills using topic modeling algorithms.

The ABSTRACT clearly presents objects, methods, and results.

The abstract could be more concise and focused. It should clearly state the main findings and contributions of the paper without going into too much detail about the background.

The introduction provides a good overview of the importance of skills in the job market. However, it could be improved by providing a clearer statement of the paper's objectives and how it contributes to the existing literature.

There are a few grammatical errors and spelling mistakes in this article.

Yes, there are.
Here is a list of some grammatical errors and spelling mistakes found in the paper:

Abstract:

"Contemporary jobseekers" should be "Contemporary job seekers."
"Since that these data" should be "Since these data."
"mainly on regard to" should be "mainly with regard to."
"Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), and BERTopic" should be followed by "are" before continuing the sentence.
Introduction:

"skills organizational" should be "organizational skills."
"industries different" should be "different industries."
"Algebraic Models like Latent Semantic Analysis, Probabilistic Models like Latent Dirichlet Allocation and Neural Models like BERTopic" should have commas after "Analysis" and "Allocation."
"rapid technological advances, demographic changes and" should have a comma after "changes."
Literature Review:

"Labor Market Intelligence recently" should be "recent Labor Market Intelligence."
"Labor Market Policy makers" should be "Labor Market Policymakers."
"analysis were made not made" should be "analysis was not made."
"Kraja & Boriçi, 2021aimed" should be "Kraja & Boriçi (2021) aimed."
"Shehu & Stringa, 2024aimed" should be "Shehu & Stringa (2024) aimed."
"Fetahu & Lekli, 2023conducted" should be "Fetahu & Lekli (2023) conducted."
"up to now Mankolli & Bushati, 2023 employs" should be "up to now, Mankolli & Bushati (2023) employ."

Challenges of Using LSA, LDA, and BERT in Albanian Language:

"several potential encountered" should be "several potential challenges encountered."
"Below, we outline several potential encountered when employing these methodologies" should be "Below, we outline several potential challenges encountered when employing these methodologies."
Conclusions:

"which should be carefully considered by every country when formulating their Labor Market Policies, because it helps to increase the efficiency of recruitment and selection of candidates" is a run-on sentence. It could be rephrased for clarity.
"This provides an advantage, because if we have can interpret easy the generated, we can find more model errors more easily." should be "This provides an advantage because if we can easily interpret the generated model, we can more easily find errors."

The study METHODS are explained clearly.

1. The paper provides a detailed explanation of the three topic modeling algorithms. However, it would be beneficial to include a more explicit discussion of how these algorithms were applied to the analysis of job advertisements in the Albanian job market.
2. The methodology section lacks a clear description of the data collection process. Details such as the source of the job advertisements, the time frame of the data collection, and the selection criteria for the advertisements should be included.

The body of the paper is clear and does not contain errors.

The paper does not provide a clear results section. It is essential to present the findings from the application of the topic modeling algorithms to the job advertisements and discuss their implications for understanding employability soft skills in the Albanian job market.
The discussion could be strengthened by comparing the results with findings from other studies or by exploring the practical implications of the results for policymakers and employers.

The CONCLUSION or summary is accurate and supported by the content.

The body of the paper provides a good overview of the topic and presents valuable information. However, there are several areas where clarity could be improved, and there are some errors present. Here are some points to consider:

Clarity:

The transition between sections and the connection between ideas could be smoother to enhance readability.
Some sentences are overly long and complex, making them difficult to understand. Breaking them into shorter sentences could improve clarity.
The methodology section lacks a detailed explanation of how the topic modeling algorithms were applied to the Albanian job market data. Providing a step-by-step

process would clarify the research approach.
Errors:

There are grammatical errors and typos throughout the paper, as previously listed. The use of technical terms and acronyms should be consistent. For example, "Latent Semantic Analysis" is sometimes referred to as "LSA" and other times as "Latent Semantic Indexing."
Some references are not properly formatted or are missing necessary information, such as the year of publication.
Content:

The paper could benefit from a more in-depth analysis of the results and a discussion of their implications for the Albanian job market.
The conclusion section should summarize the key findings more clearly and suggest directions for future research.

The list of REFERENCES is comprehensive and appropriate.

Yes, the references is fine

Please rate the TITLE of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the ABSTRACT of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the LANGUAGE of this paper.

[Poor] **1-5** [Excellent]

5

Please rate the METHODS of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the BODY of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the CONCLUSION of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the REFERENCES of this paper.

[Poor] **1-5** [Excellent]

4

Overall Recommendation!!!

Accepted, minor revision needed

Comments and Suggestions to the Author(s):

For a clearer understanding of the three algorithms, it is recommended to provide more detailed explanations, supplemented by illustrative figures and practical examples. For instance, a figure could visually represent the process of topic modeling, showing how each algorithm identifies and categorizes topics within a set of documents. Additionally, providing examples of how each algorithm has been applied in real-world scenarios, particularly in the context of analyzing employability skills, would offer valuable insights into their practical applications and effectiveness.

Reviewer C:
Recommendation: Accept Submission

The TITLE is clear and it is adequate to the content of the article.

The title effectively communicates the main focus and scope of the article. It clearly indicates that the article discusses the use of topic modeling algorithms for analyzing employability soft skills specifically within the context of the Albanian job market.

The ABSTRACT clearly presents objects, methods, and results.

Yes the abstract states that the authors aim to compare topic modeling algorithms for analyzing employability soft skills in the Albanian job market.

There are a few grammatical errors and spelling mistakes in this article.

Yes, authors should proof read their article to correct any grammatical errors or misspelling

The study METHODS are explained clearly.

Yes, the study methods are explained clearly in the article. The methods section outlines the process of data preprocessing, building the corpus, model evaluation, and comparative review of the three algorithms (LSA, LDA, and BERTopic).

The body of the paper is clear and does not contain errors.

Overall, the body of the paper is clear and well-structured, presenting a thorough analysis of the topic modeling algorithms and their application in the context of labor market intelligence.

The CONCLUSION or summary is accurate and supported by the content.

Yes, the conclusion accurately summarizes the key points discussed in the article and is supported by the content. It highlights the main findings regarding the comparison of topic modeling algorithms for analyzing employability soft skills, as well as the challenges and opportunities associated with their application in the Albanian context.

The list of REFERENCES is comprehensive and appropriate.

Some references are outdated, authors should only rely on the most recent references to establish a strong link to literature.

Please rate the TITLE of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the ABSTRACT of this paper.

[Poor] **1-5** [Excellent]

5

Please rate the LANGUAGE of this paper.

[Poor] **1-5** [Excellent]

3

Please rate the METHODS of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the BODY of this paper.

[Poor] **1-5** [Excellent]

4

Please rate the CONCLUSION of this paper.

[Poor] **1-5** [Excellent]

5

Please rate the REFERENCES of this paper.

[Poor] **1-5** [Excellent]

3

Overall Recommendation!!!

Accepted, minor revision needed

Comments and Suggestions to the Author(s):