Progress Report

I. Implementation of Neural Network

The Word2Vec architecture, which was trained on a specific dataset, was utilized to construct our neural network-based word embeddings. In particular, the Skip-Gram model was used because it excels at working with big datasets and can successfully capture the semantics of uncommon words. In order to optimize the neural network's performance and use of computer resources, it was built effectively.

II. Word Embeddings Calculation

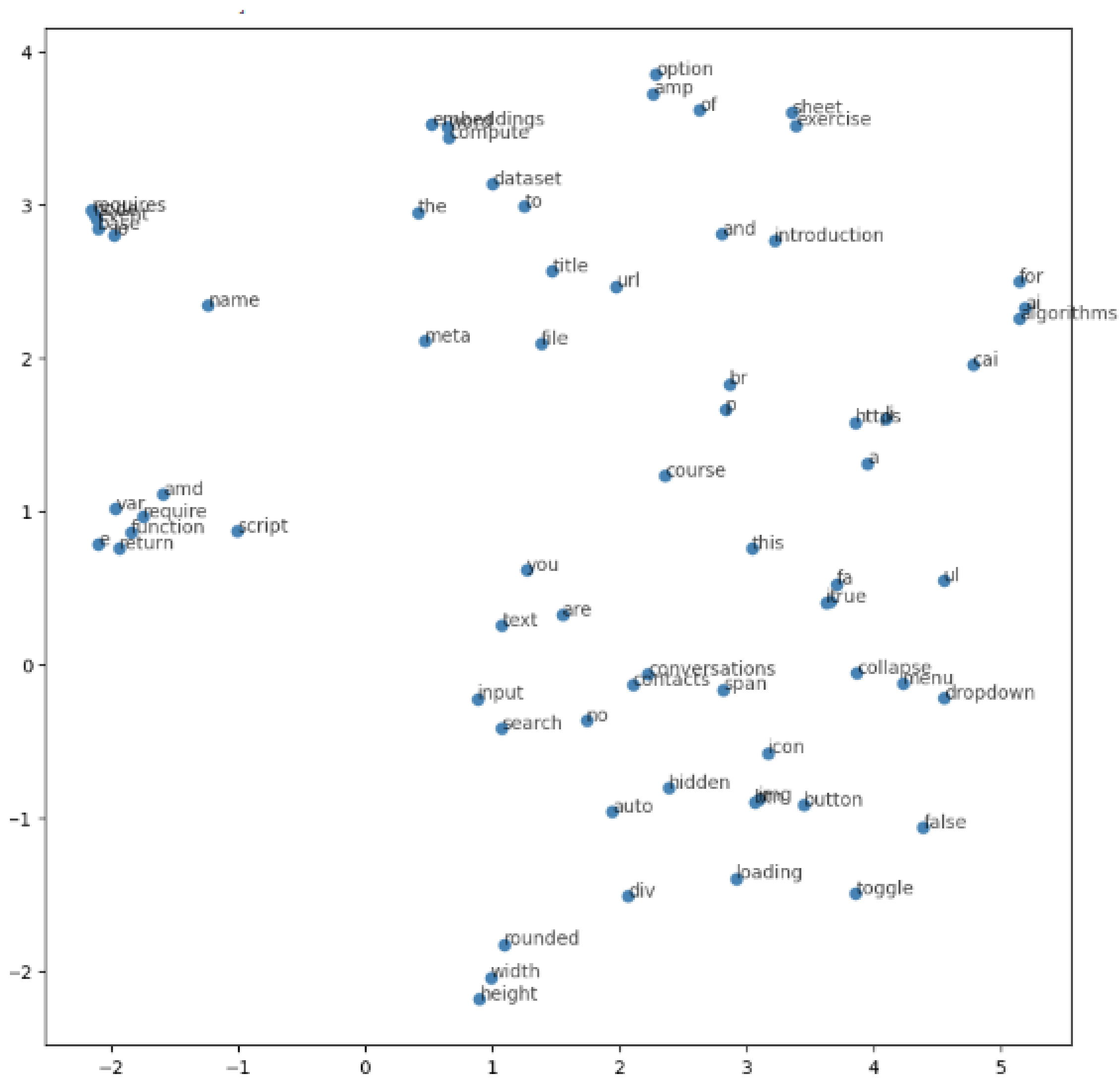
The chosen dataset was used to train the Word2Vec model, which produced word embeddings with success. These word embeddings were generated as high-dimensional vectors (with a dimensionality of 50), where each dimension represents some part of the word's meaning and words are located next to one another in this high-dimensional space if they have comparable contexts in the corpus [1].

III. Effectiveness

In this project, implementation effectiveness was given top emphasis. By calculating the program's execution time and memory usage, the efficiency was determined. The program ran in 1.325 seconds and used 245,747,712 bytes of memory, which shows that it was implemented quite effectively. The model's feasibility for bigger datasets and real-time applications is ensured by this efficiency.

IV. Visualization of Word Embeddings

We used t-SNE, a dimensionality reduction method that works particularly well for high-dimensional data visualization, for visualization [3]. It is possible to visually evaluate the discovered semantic links between words thanks to the visualization, which displays the word embeddings in a 2-dimensional space [1]. Words with comparable meanings are clustered together in this 2D space, demonstrating the efficiency of our word embeddings.



V. Next Steps

The model will be improved going forward by adjusting hyperparameters and experimenting with various neural network topologies. The objective is to raise the quality of the word embeddings, which can be judged by how well they perform on certain tasks or how closely they correspond to our intuitive understanding of the semantic links between words.

References

1. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
2. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).
3. Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(Nov), 2579-2605.