

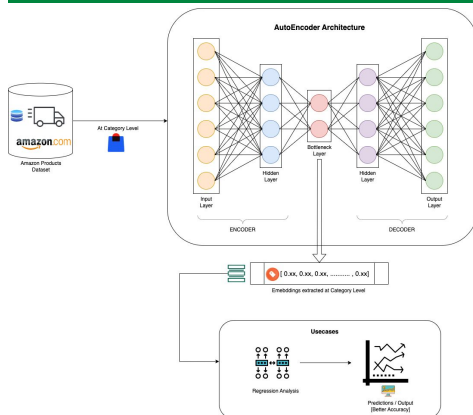
Abstract

EMBARC (Embeddings for Multilevel Product Analysis and Review Classification) is a comprehensive framework that combines the power of BERT embeddings, autoencoders and Optuna framework to generate embeddings at different levels on textual and numerical data in the dataset for analysing product categorical prices and customer reviews. This approach aims to improve the accuracy of a regression model that helps in price prediction of a product category.

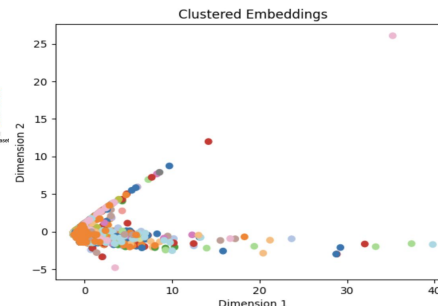
Motivation & Background

- Accurate price prediction is of utmost importance in industries such as e-commerce and retail, as it directly impacts sales and profitability.
- Leveraging embedding techniques like BERT and the Optuna, aim to create embeddings of reviews, price, rating, and number of ratings.
- The integration of these embeddings into the prediction model adds as an additional features to our use case data.
- It enhances the accuracy of our regression model for price prediction..
- This would be useful in evaluating each product category performance with respect to price..

Workflow And Implementation

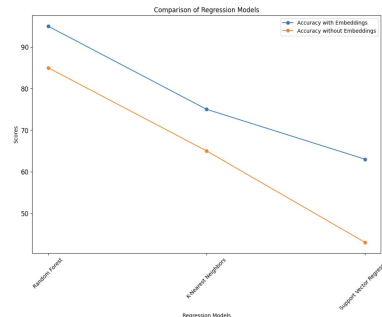


- Using autoencoders with help of Optuna and BERT, meaningful embeddings were generated to capture important features within the dataset.
- These embeddings represented the price, number of ratings, ratings, and textual content in a more condensed and informative manner, allowing for more effective analysis and modeling.
- now the regression model was computed for use case data without the embeddings to compute accuracy.



- Finally, the embeddings were added to the use case data and regression model was performed again to compute the accuracy.

Results



Regression models for price prediction with and without the embeddings and we can see the embeddings boost the accuracy of our models..

Conclusions

By leveraging BERT and Optuna, we generated comprehensive embeddings from reviews, price, ratings and number of ratings. These embeddings help to improve the price prediction of product category which could improve sales performance. The comparative analysis validates the embedding-based approach's superior accuracy and predictive power.

Future Directions

- Optimising the BERT embeddings creation using powerful GPUs
- Consider adding more features to the dataset and classifying the nature of a review using embeddings.

References

Asudani, D.S., Nagwani, N.K. & Singh, P. Impact of word embedding models on text analytics in deep learning environment: a review. Artif Intell Rev (2023). <https://doi.org/10.1007/s10462-023-10419-1>

Acknowledgements

Thank you to the Dr. Mark Albert, Dr. Ting Xiao, who were our advisors who supported us furthering this research experience through feedback and presentations that boosted our knowledge.