Literature Review

The present literature review embarks on a comprehensive exploration of the extensive research concerning multiclass classification of products based on customer reviews and multilingual sentiment analysis. Text classification is one of the most fundamental tasks of Natural Language Processing, providing many advantages across domains. One such domain is that of e-commerce, where users can check for features in items based on reviews. In this digitalized world, there is a demand for online shopping where identifying the sentiment of the reviews is important, in various languages and not just English. For this study, we conducted extensive research on the work of authors in the NLP domain and the data that would be most suitable for substantial experimentation. We compared and contrasted the different ideas presented in the research papers to understand the motivation behind the proposed methodologies. We will be discussing key themes, theories, and methodological approaches in gathering the data, the algorithms that we would be using both for product categorization and sentiment analysis, and the literature work of other authors in this domain that has shaped our understanding of this area.

(Keung et al., 2020) are the pioneers of the Multilingual Amazon Reviews Corpus (MARC) database, which we will be using for our research. This corpus consists of product reviews across different categories in seven languages, collected from 2015 to 2019. The dataset is balanced based on ratings in each language, such that each rating consists of 20% reviews in each language. The research reports baseline results for supervised classification and zero-shot classification using BERT. Since they use ratings for classification, mean absolute error (MAE) is used instead of accuracy.

There are a lot of algorithms available for developing a text classification model. These include the traditional models like Naive Bayes, K-Nearest Neighbours, Logistic Regression, and Decision Trees, as well as the recent pre-trained models like BERT, RNNs, etc. Before we can say that our model has great accuracy, we need to have a base model to compare to. (Lin et al., 2023) in their paper have recalled the importance of a simple linear classifier, which can be used as a baseline standard for comparing the performance of advanced models. According to the paper, linear SVM sometimes outperforms BERT, and the idea is that linear classifiers can serve as useful baselines and can help estimate if the advanced models are performing as expected.

The task of classifying the products into their categories can be answered by many algorithms. (Esmaeilzadeh & Taghva, 2021) examine multiple word embedding models with a focus on the Neural Network Language Model (NNLM) and BERT techniques. The authors have experimented with these two models on the IMDB review dataset, which has two categories for positive and negative reviews. NNLM had an overall precision of 0.86, while BERT showed a 6% improvement in precision for predicting positive reviews. Contrary to the idea presented in (Lin et al., 2023) the results highlight the fact that pretrained models can help achieve better performance. What we feel is that it is not always the case that linear classifiers have an edge over pre-trained classifiers. This fact, combined with the base model results, can help us understand the veracity of our approach.

However, always using pre-trained models as the starting point for developing customized classification models may not be feasible. These models are computationally very expensive as they train a large number of parameters. The base BERT model trains approximately 110 million parameters. Owing to these large numbers of parameters, the training time increases drastically. As opposed to this, the traditional methods have a lot fewer parameters, and the training time is also less. (Sagar et al., 2020) discuss well-known standard classifiers like Support Vector Machines (SVM), Naive Bayes (NB), Decision Trees (DT). Additionally, they discuss the Long Short-Term Memory technique which is a modern

method under the category of Recurrent Neural networks (RNNs). Before the models are applied, they perform text preprocessing to structure the data. An interesting observation by (Sagar et al., 2020) is that sometimes proper nouns may lose vital information when punctuation or stop words are removed. To resolve this, they use Named Entity Recognition (NER). A statement in the paper says that decision trees are the best decision-making methods. This is a debatable statement, as there are a lot of things that go into identifying the best classification method. This statement is, in fact, refuted by their results on the dataset. The results show that DT is the worst classifier when compared to the other methods, and LSTM has the best performance. LSTM has the highest accuracy, precision, and recall values. Naive Bayes ranks second and, hence, can be a good classification model.

(Haque et al., 2018) use Amazon product data for performing sentiment analysis on reviews across three categories, namely Electronics, Musical, Cellphone & Accessories. After data acquisition, they have performed various pre-processing steps such as tokenization, removal of stop-words, and POS tagging. Feature extraction methods such as bag of words, TF-IDF, and the Chi square test are important for our research to understand which of these yields the best performance when classification algorithms are applied to them. This study utilizes various classification algorithms and reports a comparative study of their performance on the datasets. Linear SVM performs the best on each dataset, giving the highest accuracy. Performance is also measured on other metrics such as precision, recall, and F-measure.

Understanding sentiments and categorizing the reviews into suitable products are not limited to just the English language. (Sharma & Parwekar, 2023) incorporates Spanish reviews for product categorization across 30 products. For the pre-processing of the data, tokenization, lemmatization, and collocation are implemented. The study employs various machine learning algorithms with MultinomialNB's performance results being inferior to those of the other algorithms. SVM with an RBF kernel performs well, with an accuracy of 72%. Optimization with Stochastic Gradient Descent improves the performance drastically and increases the accuracy to 90.10% on Spanish reviews.

In our proposed study, we intend to experiment with a range of traditional methods, such as those previously mentioned, though we remain open to exploring additional options. We will pre-process the data, including tokenization, removal of stop-words, and POS tagging, which we consider essential for our research as it will help remove noise and structure the data. Our comparative study of algorithms for product category classification and sentiment analysis will involve employing similar evaluation metrics. Furthermore, we aim to extend our analysis to encompass product classification and sentiment analysis in multiple languages, allowing us to observe how different algorithms perform across various linguistic contexts.

References:

Esmaeilzadeh, A., & Taghva, K. (2021). Text classification using neural network language model (NNLM) and Bert: An empirical comparison. *Lecture Notes in Networks and Systems*, 175–189. https://doi.org/10.1007/978-3-030-82199-9_12

Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*. https://doi.org/10.1109/icird.2018.8376299

Keung, P., Lu, Y., Szarvas, G., & Smith, N. A. (2020). The multilingual Amazon Reviews Corpus. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. https://doi.org/10.18653/v1/2020.emnlp-main.369

Lin, Y.-C., Chen, S.-A., Liu, J.-J., & Lin, C.-J. (2023a). Linear classifier: An often-forgotten baseline for text classification. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. https://doi.org/10.18653/v1/2023.acl-short.160

Sagar, S. P., Oliullah, K., Sohan, K., & Patwary, Md. F. (2020). PRCMLA: Product Review Classification Using Machine Learning Algorithms. *Advances in Intelligent Systems and Computing*, 65–75. https://doi.org/10.1007/978-981-33-4673-4_6

Sharma, P., & Parwekar, P. (2023). Multiclass classification of online reviews using NLP & Machine Learning for Non-english language. *Intelligent Human Computer Interaction*, 85–94. https://doi.org/10.1007/978-3-031-27199-1_9