Sentiment Analysis and Classification of Yelp Reviews

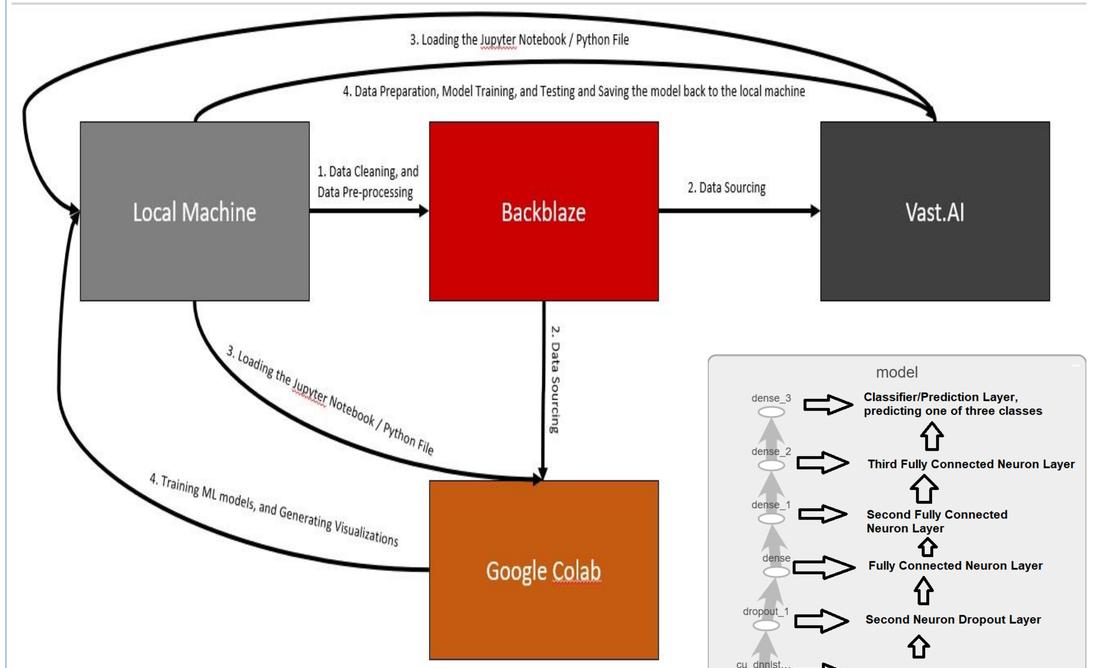
Daniel Youssef | Informatics and Computer Science | Prof. Gerard McKee

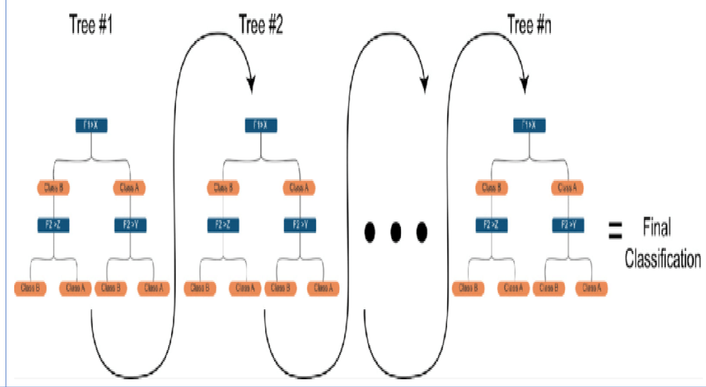
Abstract

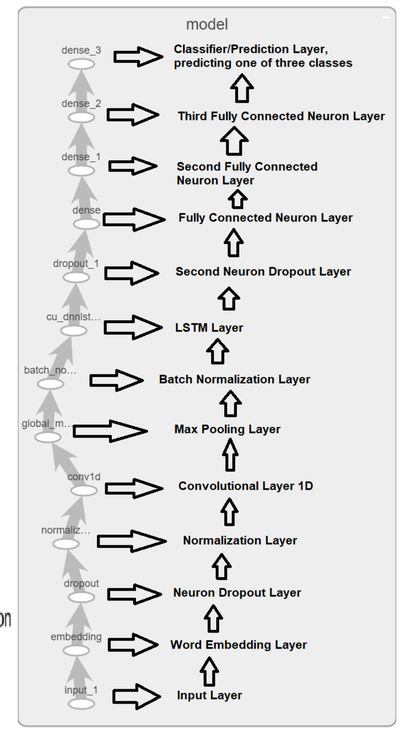
In the field of Natural Language Processing, one of the problems that has varying degrees of effective solutions is the ability to perform Sentiment Analysis on textual data, and to retrieve an AI-based solution that is able to predict what the textual data conveys. In order to attempt to reach a reliable solution to this problem, a large body of research, as well as related work was analysed, and summarized. As a result of this, a proposed solution was implemented, utilizing a dataset containing 6990280 reviews, and a novel technique for resampling data to further improve upon prediction scores for underrepresented classes. The proposed models utilized an XGBoost architecture, and an LSTM architecture, to achieve accuracy scores of 85%, and 77% respectively.

Methodology

Two models were proposed, using XGBoost [1], and LSTM [2] architectures to create two models that can detect sentiment in text. In addition, these two architectures were used to detect sarcasm in text as well. The pipeline visualized below explains the process of that was used to train these models and generate their visualizations [3].

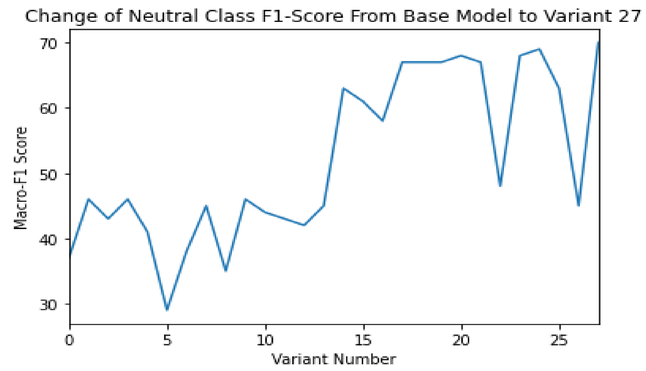






Results

Model	Score
XGB SMOTE	86%
LSTM V27	77%
XGB CUST	71%



Conclusions

- CUST improves predictions of the Neutral Class in Sentiment Analysis by over 20%.
- XGB and LSTM are two proposed models depending on use case, both achieve state-of-the-art performance on Main Yelp Review Dataset.
- Sentiment Analysis, and Sarcasm Detection done using ML, and DL techniques.

References

[1] M. Seyfioğlu and M. Demirezen, “A hierarchical approach for sentiment analysis and categorization of Turkish written Customer Relationship Management Data,” Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, 2017. doi:10.15439/2017f204

[2] Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, and Mounika Bagadi, Mounika Belusonti, “Text based sentiment analysis using LSTM,” International Journal of Engineering Research and, vol. V9, no. 05, 2020. doi:10.17577/ijertv9is050290

[3] J. Abualdenien and A. Borrmann, “Ensemble-learning approach for the classification of levels of geometry (log) of building elements,” *Advanced Engineering Informatics*, vol. 51, p. 101497, 2022. doi:10.1016/j.aei.2021.101497