2021-2022 M.Sc. in Data Science and Analytics

CUSTOMER REVIEW CLASSIFICATION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Nikhil Sunil Supervisor: Dr. Farid Shirazi

Objective:

The objective of this research is to predict the sentiment of customer reviews using machine learning models with a vectorizer and deep learning techniques and select the best modelling technique based on their performance metrics.

Background

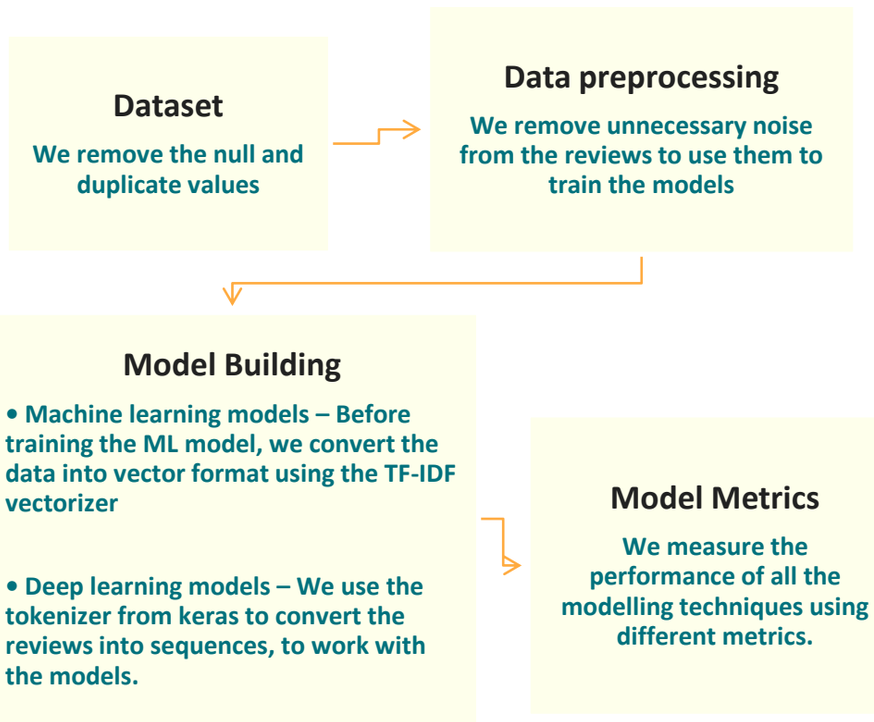
Customer reviews are pieces of feedback that are given to businesses and Retailers based on the customer's experience, with their organization. They are used by companies to improve upon their existing service or the product they are selling. In an e-commerce driven world, where people have no physical access to the goods they wish to purchase, many customers will turn to online reviews to get an opinion on what to buy.

E-Commerce is really changing the way in which, people buy products and services. To move up in the corporate world, many businesses are finding different ways to increase their customer count by looking at past data. Companies can use customer reviews on their products to promote their high selling products and gain a competitive edge in the market. We focus on classifying the sentiment of customer reviews with the use of different machine learning and deep learning modelling techniques.

Methodology

1. **Dataset:** We first address the missing and duplicate values in the dataset and identify the features we need for the training the model i.e. the reviews and recommendation index.
2. **Text processing:** For processing the reviews, we have used the following techniques to remove unnecessary noise.

- Removing punctuations
- Tokenization
- Removing numbers
- Filtering stop words
- Lemmatization



3. **Model building**: We train 5 machine learning classifiers and 3 recurrent neural networks to classify the sentiment of customer reviews . We vectorize the processed reviews for model training using the TF-IDF vectorizer for the machine learning models and the Keras Tokenizer and sequence function for the deep learning models.

4. **Metrics:** The Metrics we will be using to measure the model performance are as follows.

- Confusion Matrix
- AU-ROC curve:
- Classification report

Results

Exploratory Data

Based on the exploratory analysis, we found that the proportion of classes were unevenly balanced, which meant that the dataset class ratio needed to be upscaled.

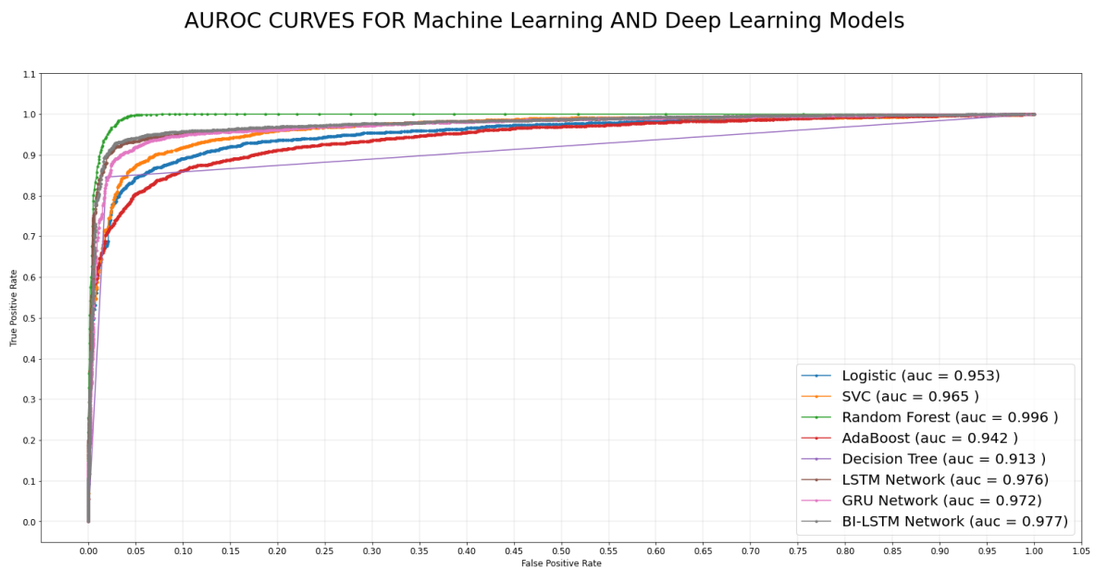
Classification Report:

The random forest model edges out the rest of the models with higher accuracy and a better score when predicting the test data.

	Train_Score	Test_Score	Precision_Score	Recall_Score	F1_Score	accuracy
Random Forest Tree	1.000000	0.971097	0.979514	0.967098	0.973266	0.971097
BI-LSTM	0.990023	0.947330	0.964494	0.937702	0.950909	0.947330
LSTM	0.994645	0.941755	0.976129	0.915318	0.944746	0.941755
GRU	0.970435	0.933979	0.946057	0.931769	0.938859	0.933979
Support Vector Machine	0.940760	0.913146	0.938626	0.899137	0.918457	0.913146
Decision Tree	1.000000	0.907424	0.981534	0.845739	0.908590	0.907424
logisitc regression	0.918018	0.893633	0.908071	0.895092	0.901535	0.893633
AdaBoost	0.901181	0.870745	0.876632	0.887271	0.881919	0.870745

AUROC Curve:

Random Forest classifier has once again achieved the best performance as it has the highest AUC value of 99.6% and it's ROC curve is higher than that of other models. LSTM and BI-LSTM are very close in terms of AUC values (LSTM AUC= 0.976, BI-LSTM AUC = 0.977) and their ROC curves are almost on a similar trajectory. The GRU model sees a slight dip with a AUC value of 0.978 and similar RUC curve trajectory. The 3 bottom ROC curves are produced by the remaining ML models (SVC, Logistic Regression, AdaBoost and Decision Tree) which fared worse than the RNN models with low AUC scores and no upward incline of ROC curves.



Conclusions

Among all the models, The Random Forest Classifier with TF-IDF vectorizer seemed to predict sentiments with the highest accuracy and better AU-ROC curve than the rest. Based on overall performance, it can be concluded that the Deep learning models have a better overall score than the machine learning models.

For future works, we could consider the option of adding emojis to the dataset, which could help the models predict sentiments better. Reviews may contain many stop words that may remain even after processing, which can be classified as positive sentiment, even though the overall sentiment is a negative one. We can also look to further fine tune the hyperparameters of the deep learning networks to perform better, predicting sentiments.