4222 – SURYA GROUP OF INSTITUTIONS

VIKRAVANDI

SENTIMENT ANALYSIS FOR MARKETING

PREPARED BY

S.Masilamani

422221106304

ECE DEPARTMENT

SENTIMENT ANALYSIS FOR MARKETING

AI_Phase 2

MACHINE LEARING

INTRODUCTOIN:

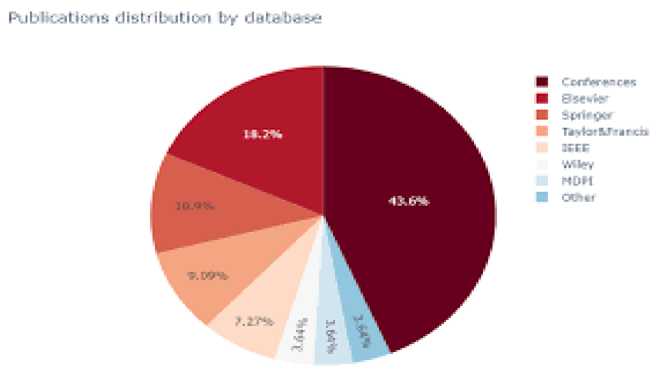
Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data. The field of study known as machine learning is concerned with the question of how to construct computer programs that automatically improve Comparison

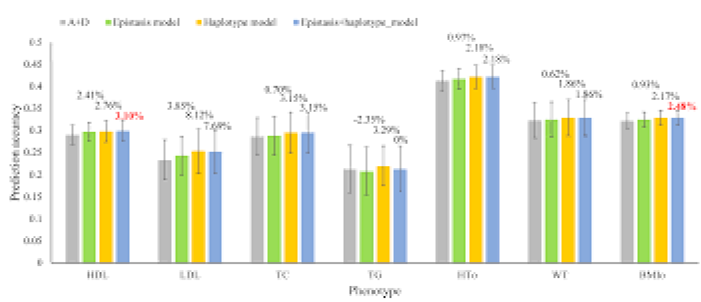
DEFINITION OF LEARNING:

A computer program is said to learn from experience E with respect to some class of task T and performance measure P,it its performance attasks T, as measured by P,improves with experience E. To answer the question of how hard the two tasks are, we can compare our system's performance against that of humans. We conducted an scaled-down version of the experiment where we had humans attempt the same two classification task as our models. Performance at the human level is often considered the target goal in sentiment analysis. experience.

HUMAN PREDICTION:

.Similar to sales forecasting, are based on datasets from past prices, volatility indices, and fundamental indicators. Beginners can start small with a project like this and use stock-market datasets to create predictions over the next few months. It's a great way to become familiar with creating predictions based on massive datasets.one might ask what is the difficulty of our two task and what level of accuracy would be considered successful. Answer





Other Attempt:

In addition to what we used in our final model, we had other work that taught us more about extracting emotion from EP. 3 For the max label task, due to the unbalanced distribution of categories we used a balanced human testing set instead of a random subset of the original testing set. Note that this is a harder problem for our SVM classifier since it was trained on an unbalanced training set. As a result the numbers reported here are lower than the ones reported in Results

. Figure 4: Human comparison for max label.

params = list(model.named_parameters()) print("The BERT model has {:} different named parameters.".format(len(params)) print("==== Embedding Layer ====") for p in params[0:5: print(".format(p[0], str(tuple(p[1].size())))) print("==== First Transformers ====") for p in params[5:21]: print("{:12}".format(p[0], str(tuple(p[1].size()))) print("==== Output Layer ====") for p in params[-4:]: print("{:12}".format(p[0], str(tuple(p[1].size()))))

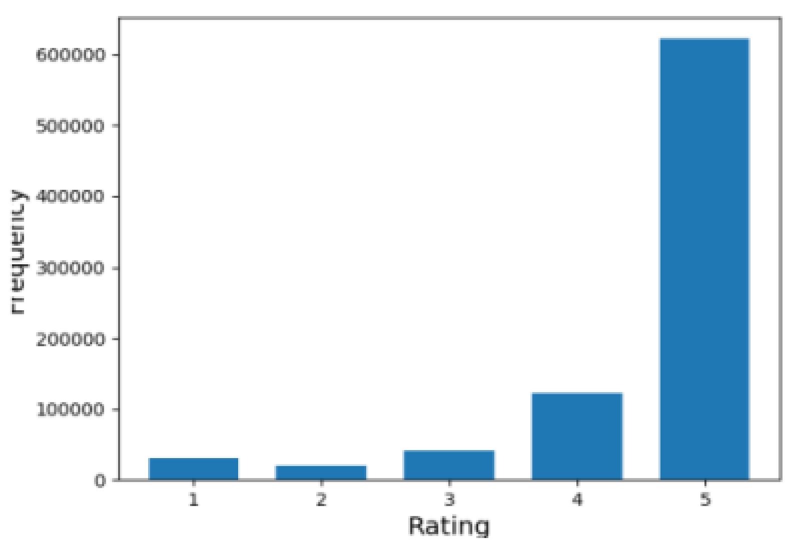
Furthermore

he user generated information may also contain unnecessary whitespaces at the beginning, in between or at the end of the tweets, special characters like punctuation and repetition of characters. First, all extra white space was removed using the build in function available in Python. Secondly, all the meaningless and unnecessary special characters from the tweets were eliminated (Hemalatha et al. 2012).

These characters include: \ | [ ] ;: {} - + ( ) < > ? ! @ # % *, and a few more. Neither do these characters have specific and special meaning, nor do they explain if these characters are used for positivity or negativity, hence; removing them is the best option • Non-standard (slang) to standard word mapping • PoS tagging • Tagging Stopword removal • positive/negative word

Furthermore, the user generated information may also contain unnecessary whitespaces at the beginning, in between or at the end of the tweets, special characters like punctuation and repetition of characters. First, all extra white space was removed using the build in function available in Python. Secondly, all the meaningless and unnecessary special characters from the tweets were eliminated (Hemalatha et al. 2012). These characters include: \ | [ ] ;: {} - + ( ) < > ? ! @ # % *, and a few more. Neither do these characters have specific and special meaning, nor do they explain if these characters are used for positivity or negativity, hence; removing them is the best option • Non-standard (slang) to standard word mapping • PoS tagging • Tagging Stopword removal • positive/negative words

```
import numpy as np print('---Train---')
 print('input: ', train_input.shape)
 print('label: ', train_labels.shape)
 print('mask: ', np.array(train_mask).shape)
 print('---Validation---')
print('input: ', validation_input.shape)
 print('label: ', validation_labels.shape)
 print('mask: ', np.array(validation_mask).shape)
print('---Test---')
print('input: ', test_input.shape)
```



CONCLUSION:

Sentiment analysis deals with the classification of texts based on the sentiments they contain. This article focuses on a typical sentiment analysis model consisting of three core steps, namely data preparation, review analysis and sentiment classification, and describes representative techniques involved in those steps.