



4222 – SURYA GROUP OF INSTITUTIONS

VIKRAVANDI

SENTIMENT ANALYSIS FOR MARKETING

Prepared by,

S.DHANUSH

REG. NO: 42221106018

ECE 3rd year

PHASE 2 :INNOVATION

MACHINE LEARNING

Introduction

Sentiment analysis is a type of machine learning tool. Machine learning works with natural language processing to make up the core building blocks of the sentiment analysis process. Surprisingly, one model that performs particularly well on sentiment analysis tasks is the convolutional neural network, which is more commonly used in computer vision models.

Machine learning approach

Machine learning involves showing a large volume of data to a machine so that it can learn and make predictions, find patterns, or classify data. The three machine learning types are supervised, unsupervised, and reinforcement learning.

- *Lexicon-based Methods.*
- *Automated/Machine Learning Methods.*
- *Hybrid approach*

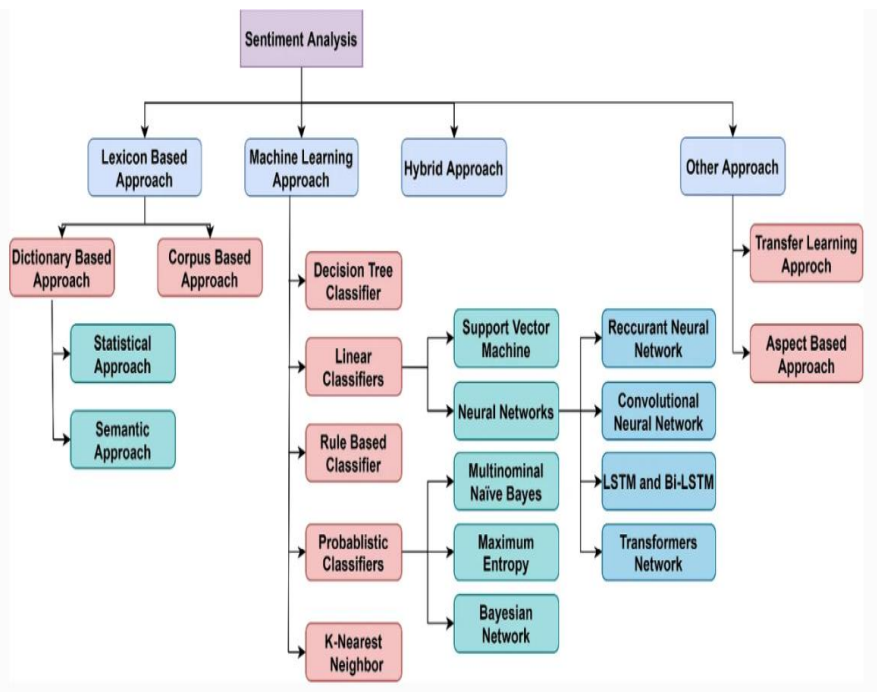


Fig:1 Sentiment Analysis

Machine learning algorithm

- ❖ *Supervised*
- ❖ *semi-supervised*
- ❖ *unsupervised*
- ❖ *reinforcement.*

Comparison to Human Prediction:

One might ask what is the difficulty of our two tasks and what level of accuracy would be considered successful. To answer the question of how hard the two tasks are, we can compare our system's performance against that of humans. We conducted a 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.

Best dataset for sentiment analysis

- ✚ *. Amazon Review Data.*
- ✚ *Stanford Sentiment Treebank.*
- ✚ *Financial Phrasebank.*
- ✚ *Webis-CLS-10 Dataset.*
- ✚ *CMU Multimodal Opinion Sentiment and Emotion Intensity.*
- ✚ *Yelp Polarity Reviews.*
- ✚ *WordStat Sentiment Dictionary.*
- ✚ *Sentiment Lexicons For 81 Languages*

If you have ever looked up sentiment analysis online, chances are you've come across a project analyzing Twitter feeds. This is an excellent project because there are millions of public tweets on Twitter every day, as well as being housed by various APIs that work to collect content. You can use a Twitter crawler or an API source to build a dataset of a portion of these tweets and analyze them.

A great project is to build a tweet dataset on a specific topic or hashtag, and categorize each tweet's sentiment as positive or negative, with the ultimate goal of forming an aggregated sentiment on the topic as a whole.

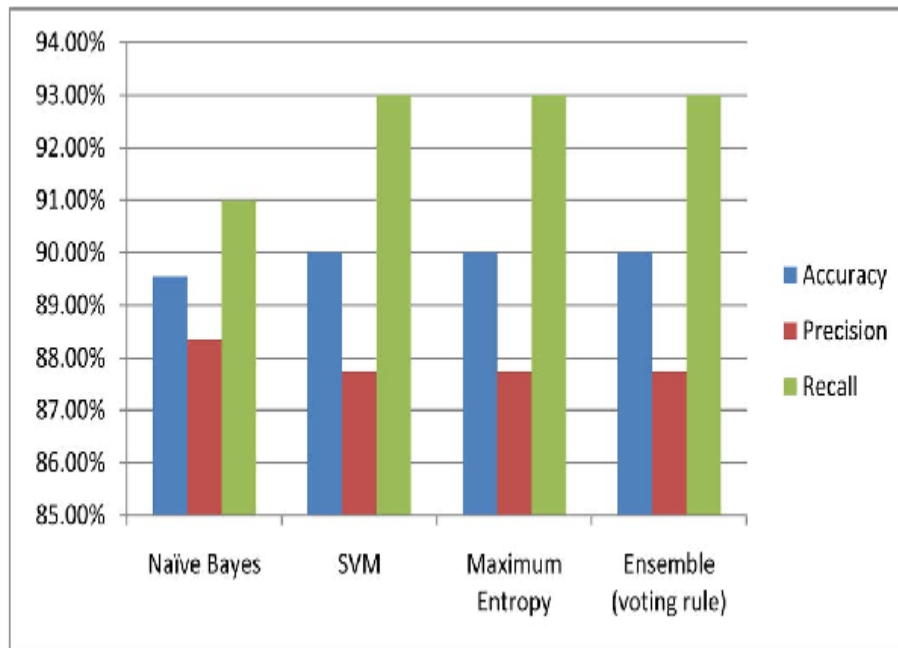


Fig. 2. Performance of Different classifiers in Twitter Sentiment Analysis

Book Sentiment Analysis

As a book lover, I always look for ways to leverage what I already love to learn new things. So, if you like books and novels, you can build a sentiment analyzer for your favorite book and learn all the basics of sentiment analysis as you do so. You can do that by downloading your favorite book as a pdf and then processing and manipulating the text. You can find a similar project using R [here](#).



Fig:3 It is represent the book sentiment analysis

Data Cleaning

Data cleaning in sentiment analysis is the process of removing redundant and incorrect values in data that is meant for analysis. This is a necessary step in the sentiment analysis process, whatever the business requirement may be - whether customer experience analysis, employee satisfaction analytics, or brand experience insights. Removing all the unnecessary data items that do not belong in your dataset is an essential part of sentiment analysis data preparation, without which the insights you receive will be inaccurate and cannot be relied on.

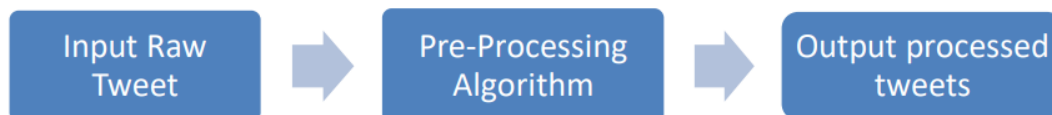


fig: 4 Data cleaning process

Random Forest Algorithm

We upload the dataset (.csv file) by using the pandas API in python and we apply the TfidfVectorizer. After that we train the dataset and save the model file in the '.sav' file format. We store the fit model in .sav format for future testing purposes. Random Forest algorithm is a supervised algorithm for both

classification and regression algorithm. It computes the decision based on the highest scores of

multiple decision trees. In the following figure, we can see the implementation code of the training the dataset using Random Forest algorithm.

```
# Visualising the Random Forest Regression results

# arrange for creating a range of values
# from min value of x to max
# value of x with a difference of 0.01
# between two consecutive values
X_grid = np.arange(min(x), max(x), 0.01)

# reshape for reshaping the data
# into a len(X_grid)*1 array,
# i.e. to make a column out of the X_grid value
X_grid = X_grid.reshape((len(X_grid), 1))

# Scatter plot for original data
plt.scatter(x, y, color='blue')

# plot predicted data
plt.plot(X_grid, regressor.predict(X_grid),
         color='green')
plt.title('Random Forest Regression')
plt.xlabel('Position level')
plt.ylabel('Salary')
plt.show()
```

The steps for trading with sentiment analysis

Following are the steps for trading with sentiment analysis

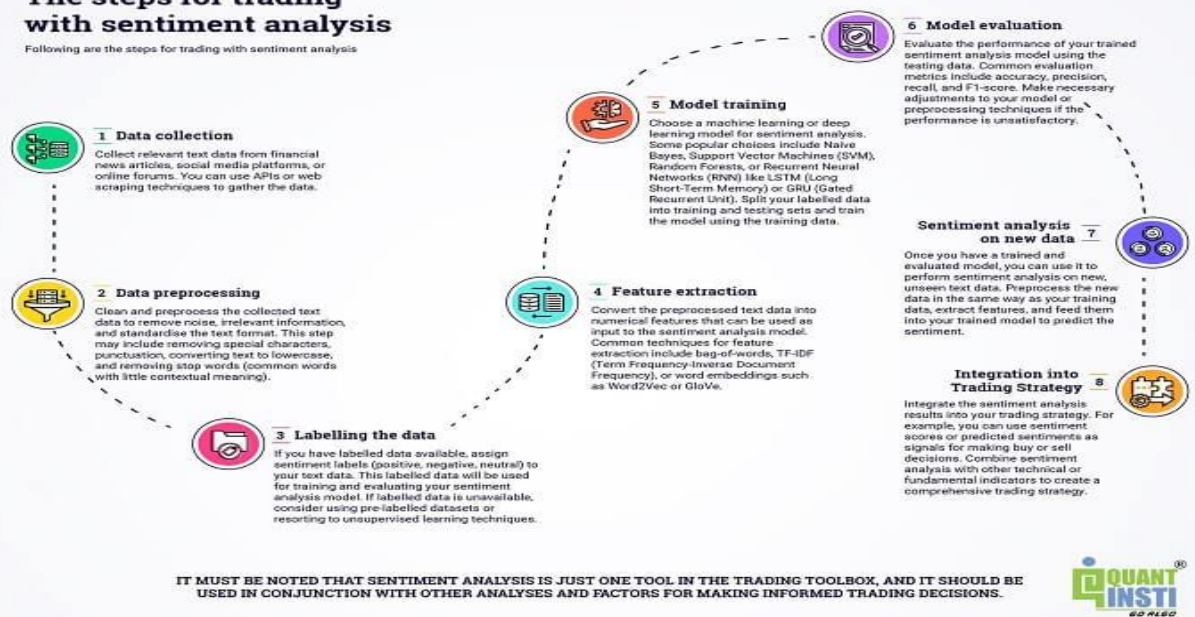


Fig:5 Steps for trading with sentiment analysis

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.

```
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("{:<60} {:>12}".format(p[0], str(tuple(p[1].size()))))
print("==== Output Layer ====")
for p in params[-4:]:
    print("{:<60} {:>12}".format(p[0], str(tuple(p[1].size()))))
Data Cleaning
params = list(model.named_parameters())
print("The BERT model has {} different named parameters.".format(len(params)))
```


Conclusion

Social Media Sentiment Analysis can be defined as extracting sentiment or opinions out of the content posted on various social media platforms such as Twitter, Facebook, etc. This analysis can help businesses understand in which areas their products are working well and in which areas they are working poorly. Sentiment analysis using machine learning can help any business analyze public opinion, improve customer support, and automate tasks with fast turnarounds. Not only saving you time, but also money.