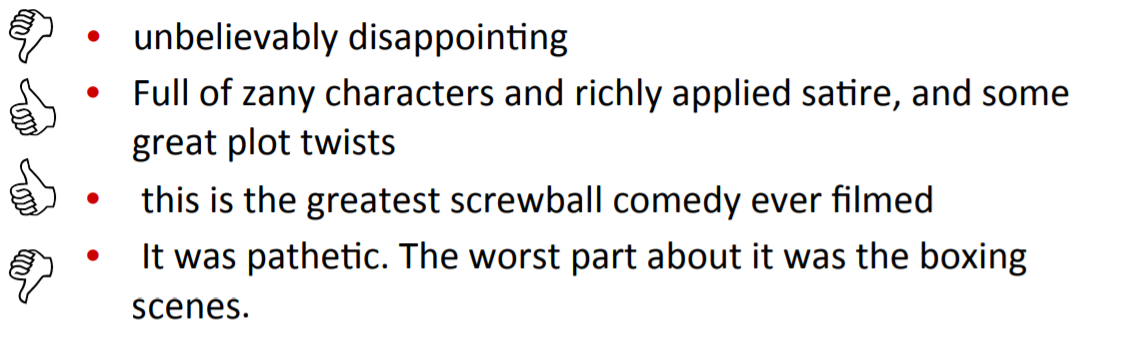
# Sentiment Analysis

Sentiment analysis is the process of understanding the opinion of an author about a subject.

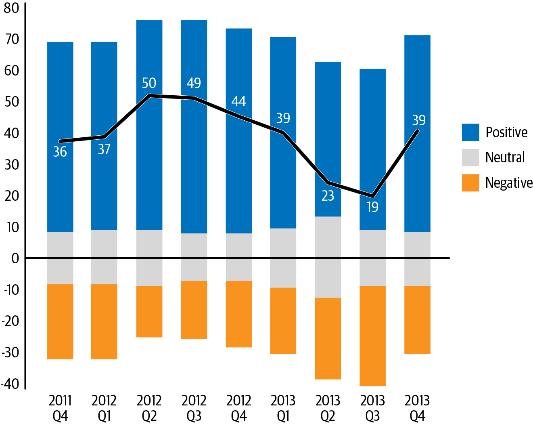
**Defination**- Sentiment analysis is also popularly known as opinion analysis or opinion mining. The key idea is to use techniques from text analytics, NLP, machine learning, and linguistics to extract important information or data points from unstructured text.

Sentiment analysis is perhaps one of the most popular applications of natural language processing and text analytics, with a vast number of websites, books, and tutorials on this subject. Sentiment analysis seems to work best on subjective text, where people express opinions, feelings, and their mood

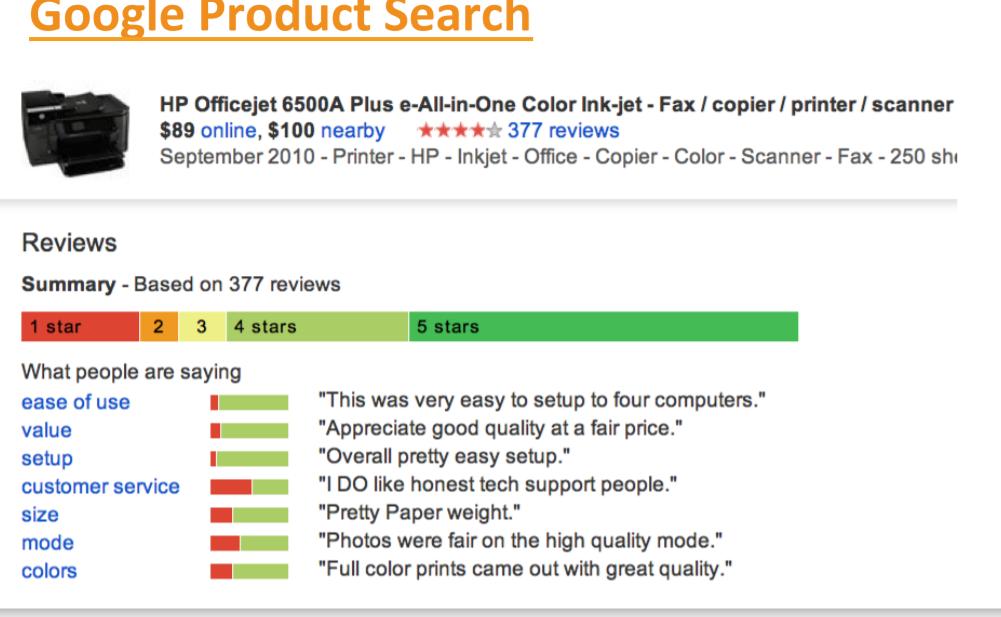
Movie Reviews



Twitter Reviews

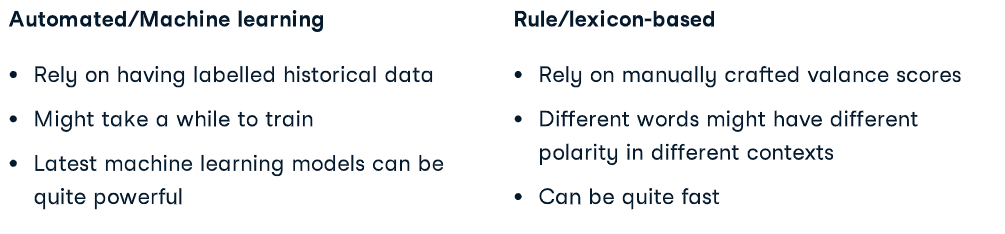


Product Reviews



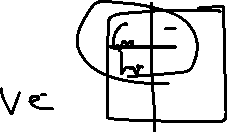
# Types of Sentiment Analysis

1. Lexicon based – Rule based
2. Machine Learning based – Learning Model

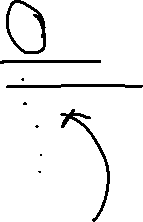




## **Lexicon-based sentiment analysis**.



*tweet that says, “The new James Bond movie is great!” is clearly expressing a positive sentiment, whereas a tweet that says, “I would never visit this restaurant again, horrible place!!” has a negative sentiment.*



*create lists of positive and negative words in English—i.e., words that have a positive or negative sentiment. We then compare the usage of positive versus negative words in the input tweet and make a prediction based on this information*

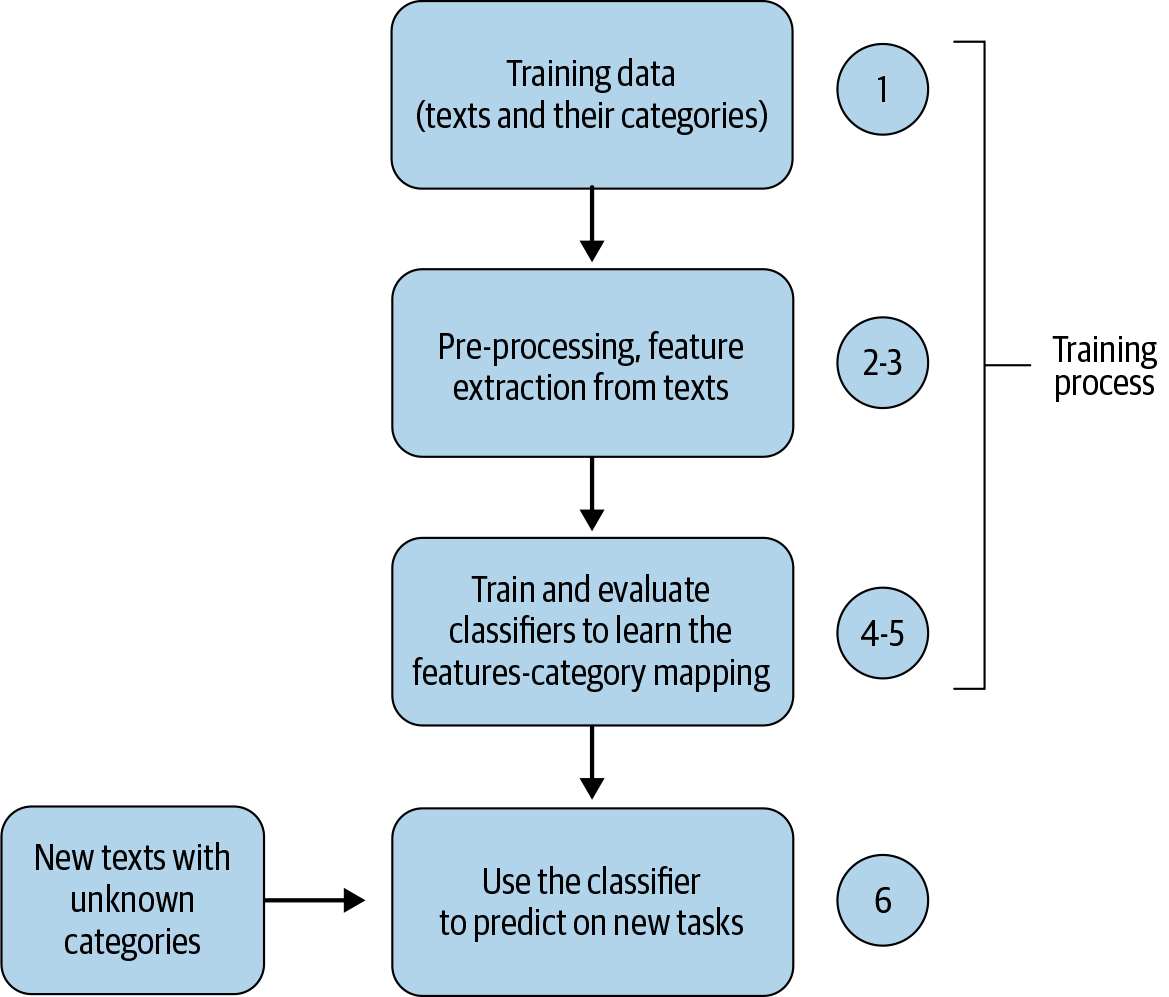
1. No “learning”



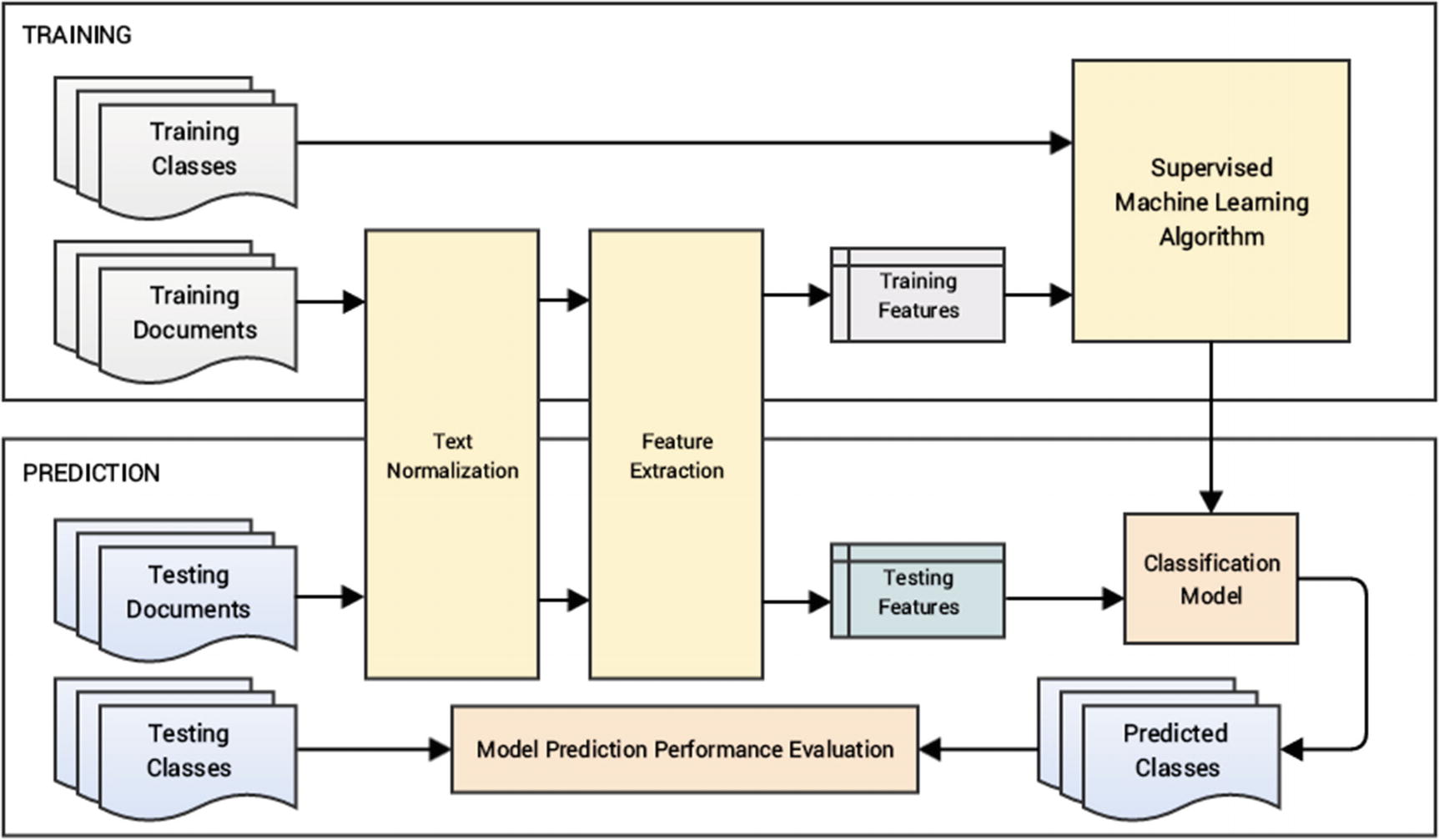
1. based on a set of heuristics or rules and custom-built resources such as dictionaries of words with sentiment



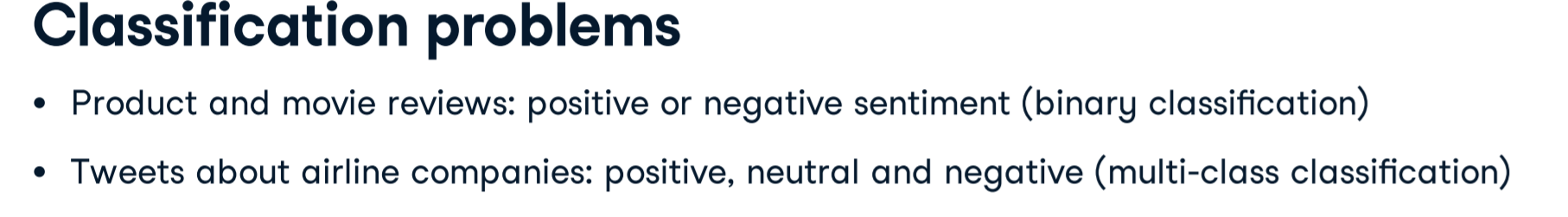
## **Learning Model based**



Sentiment Analysis - Blueprint

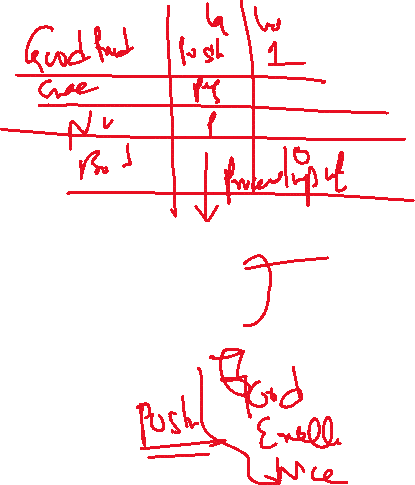




**Training Algorithms for Sentiment Analysis**

**Machine learning Based**

1. **Naïve Based- Bayes Theorem- Conditional Probability**
2. **SVM**
3. **Random forest**
4. **Logistic regression**



**Amazon review – most recent**

**Good product, excellent book,nice features**



**Worst service, losing battery soon, not working properly**

**Data- labeled**

|  |  |  |
| --- | --- | --- |
|  | **Labeling** | **Label id** |
| **Good product** | **Positive** | **1** |
| **excellent book** | **Positive** | **1** |
| **Worst service** | **Negative** | **0** |
| **losing battery soon** | **Negative** | **0** |
| **nice features** | **Positive** | **1** |
| **not working properly** | **Negative** | **0** |

**Preprocessing**

Punctuation,lower cases,stop words

**Feature vector**

BOW, TFIDF or Word Embeddings

**Training Algorithm**

1. **Machine Learning** 
   1. **Naïve Based- Bayes Theorem- Conditional Probability**
   2. **SVM**
   3. **Random forest**
   4. **Logistic regression**
2. **Deep Learning – CNN, RNN**
3. **Transfer Learning- BERT**

**Evaluation**

* **F1, recall, accuracy, probability ,confusion matrix**

**Predict -**

* **Test- good ,bad- positive ,negative- sentiment analysis –**



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Labeling** | **Label id** | **Naïve Bayes- algorithm** | **Logistic Regression algorithm** | **SVM** | **CNN** | **RNN** |
| **Good product** | **Positive** | **1** | **80%** |  |  |  |  |
| **excellent book** | **Positive** | **1** | **85%** |  |  |  |  |
| **Worst service** | **Negative** | **0** | **20%** |  |  |  |  |
| **losing battery soon** | **Negative** | **0** |  |  |  |  |  |
| **nice features** | **Positive** | **1** |  |  |  |  |  |
| **not working properly** | **Negative** | **0** |  |  |  |  |  |

