

# Solving Sentiment Analysis with a Rule-based Approach

---



**Vitthal Srinivasan**

CO-FOUNDER, LOONYCORN

[www.loonycorn.com](http://www.loonycorn.com)

# Overview

**Introduce rule-based approaches**

**Demonstrate a simplistic rule-based approach**

**Refine that approach to make it more robust**

**Apply sentiment lexicons**

# Building Is Hard, Using Is Easy



**Builder**

Building a sophisticated  
sentiment analysis system is **hard**



**User**

Using a sophisticated sentiment  
analysis system is **easy**

# A Simplistic Rule-based Approach to Polarity Detection

---

# Sentiment Analysis as Binary Classification



Attempt to build a simple rule-based classifier to classify a text fragment

# A Simple Rule-based Classifier



**Split text fragment  
into words**



**Calculate polarity of  
individual words**



**Aggregate word  
polarities**



**Split text fragment  
into words**

**Split documents into paragraphs**

**Split paragraphs into sentences**

**Split sentences into words**

**Simple library functions available**



**Split text fragment  
into words**

“This is the worst restaurant in the metropolis, by a long way”



This
is
the
worst
restaurant
in
the
metropolis
by
a
long
way





Calculate polarity of  
individual words

Tag each word as positive or negative

Ignore neutral words entirely

Requires use of a **sentiment lexicon**

Like a dictionary, but for **polarity** lookup

Ideally should also provide **intensity**

“This is the worst restaurant in the metropolis, by a long way”

This  
is  
the  
worst  
restaurant  
in  
the  
metropolis  
by  
a  
long  
way

This  
is  
the  
worst  
restaurant  
in  
the  
metropolis  
by  
a  
long  
way



[illegible]

[illegible]

[illegible]

[illegible]



## Calculate polarity of individual words

**“This is the worst restaurant in the metropolis, by a long way”**



This  
is  
the  
**worst**  
restaurant  
in  
the  
metropolis  
by  
a  
long  
way

[illegible]

“This is the worst restaurant in the metropolis, by a long way”



Calculate polarity of individual words

This
is
the
worst
restaurant
in
the
metropolis
by
a
long
way



Neutral
Neutral
Neutral
Negative
Neutral



“This is the worst restaurant in the metropolis, by a long way”

The diagram illustrates a sequence-to-sequence model. It starts with an input sentence: "This is the worst restaurant in the metropolis by a long way". This sentence is fed into a neural network, represented by a stack of blue boxes. The output of the network is a sequence of words: "This", "is", "the", "worst", "restaurant", "in", "the", "metropolis", "by", "a", "long", "way". An orange arrow points from the input sentence to the first word, "This". Another orange arrow points from the last word, "way", to the final word in the output sequence, "way".

This
is
the
worst
restaurant
in
the
metropolis
by
a
long
way







**Aggregate word  
polarities**

**More positive words? Fragment is positive**

**Else negative**

**If intensity available, sum rather than count**



Aggregate word  
polarities

“This is the worst restaurant in the  
metropolis, by a long way”

This	Neutral
is	Neutral
the	Neutral
worst	Negative
restaurant	Neutral
in	Neutral
the	Neutral
metropolis	Neutral
by	Neutral
a	Neutral
long	Neutral
way	Neutral



**Aggregate word  
polarities**

“This is the worst restaurant in the  
metropolis, by a long way”



**Negative**

# Sentiment Analysis as Binary Classification



This illustrated the basic idea of a rule-based polarity detection system

# Sentiment Analysis as Binary Classification



But this is a very simplistic system, riddled with flaws

# Limitations of a Simplistic Rule-based Approach

---



# A Simplistic Rule-based Classifier



# A Simplistic Rule-based Classifier



# A Simplistic Rule-based Classifier

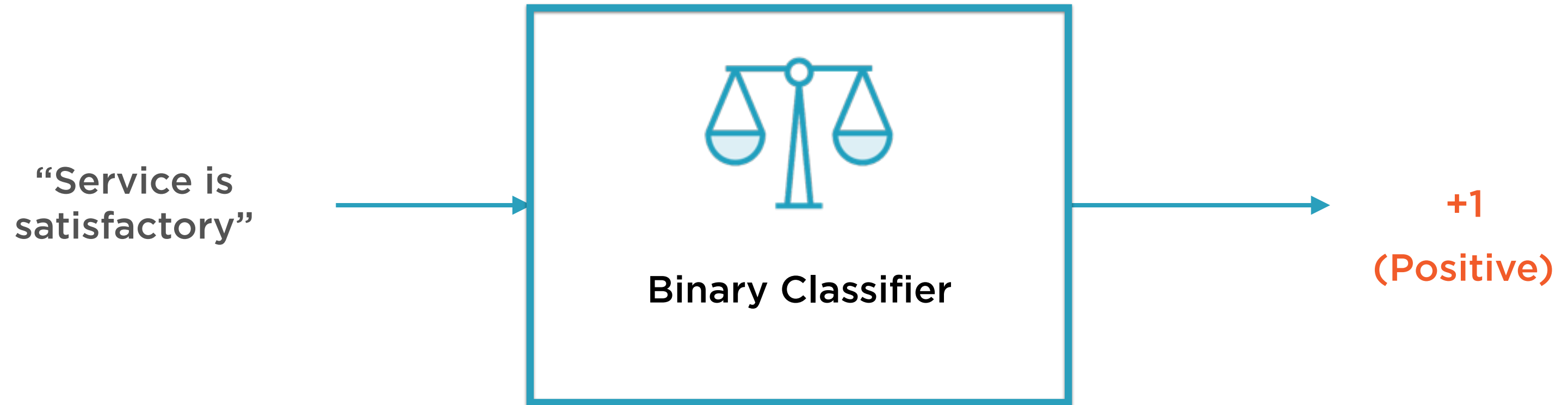
“This is **not** the worst restaurant in the metropolis, **not** by a long way”



**Binary Classifier**



# A Simplistic Rule-based Classifier



# A Simplistic Rule-based Classifier

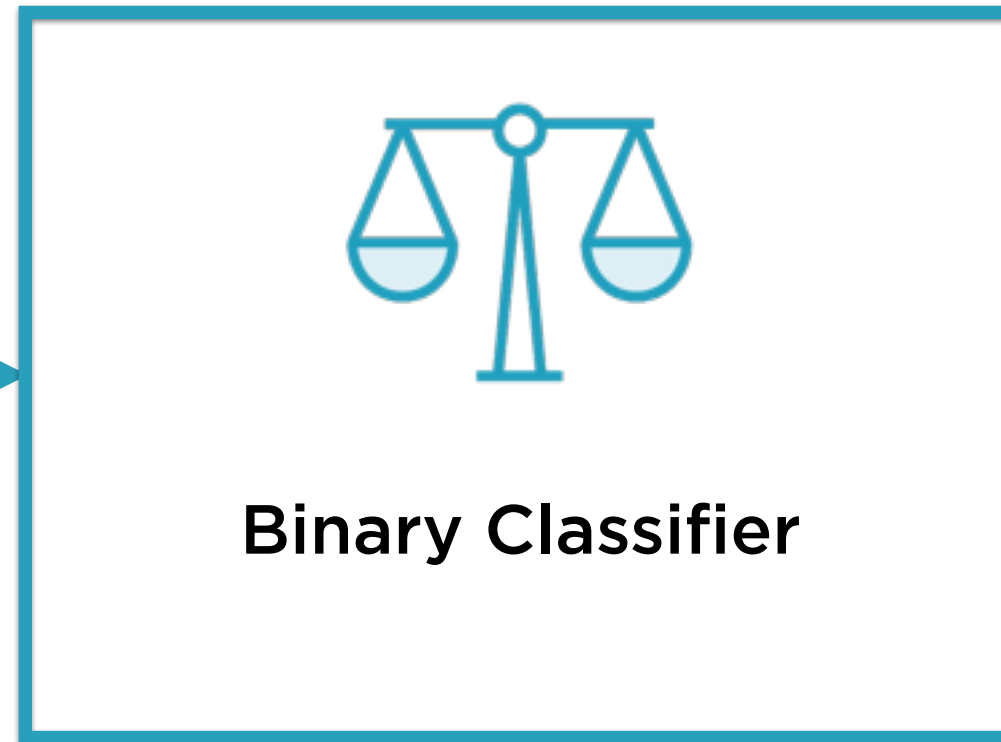


# A Simplistic Rule-based Classifier



# A Simplistic Rule-based Classifier

“Service is amazing  
**although** the food is  
terrible”



# A More Realistic Rule-based Algorithm

---



# Limitations of a Simplistic Approach



## Intensity

Valency, boosters,  
punctuation,  
capitalisation



## Reversal

Negation, contrasting  
conjunctions and  
adverbs



## Context

Different meanings in  
different contexts



Polarity alone loses intensity information

Augment sentiment lexicon with **valence scores**

Good sentiment lexicons have fine-grained valence scores for words



Good	Positive
Great	Positive
Fine	Positive
Amazing	Positive

**This sentiment lexicon lacks valence scores**

**Words have polarity, but intensity is not captured**



Good	+1.8
Great	+2.3
Fine	+2.0
Amazing	+2.9

**This sentiment lexicon includes valence scores**

**Polarity and intensity are both captured**



Intensity modulated by **boosters**

“really”, “so”, “such”

“The food is **really** good”

“The polenta is **so** good”

“That was **such** a good cake”



Intensity also modulated by  
**punctuation**

“?”, “!”

“The food is good!!!!”

“Chili-flavored icecream???”



Intensity also modulated by  
**capitalisation**

“The food is **GOOD**”



## Reversal

Negation, contrasting  
conjunctions and  
adverbs

Polarity is flipped by **negation**

“The food is **not** good”

“This is **not** the **worst** restaurant in the metropolis”





## Reversal

Negation, contrasting  
conjunctions and  
adverbs

Polarity is subtly influenced by  
**contrast**

“The food is great **but** the service is  
not”

“The noise level is annoying. **However,**  
the energy more than makes up for it”



**Context**

Different meanings in  
different contexts

**“Large”, “Small”, “High”, “Low”**

**Problematic ambiguous adjectives**

**Rule-based analysers often struggle  
with context**

**ML-based systems tend to do better if  
‘trained’ with the right data**

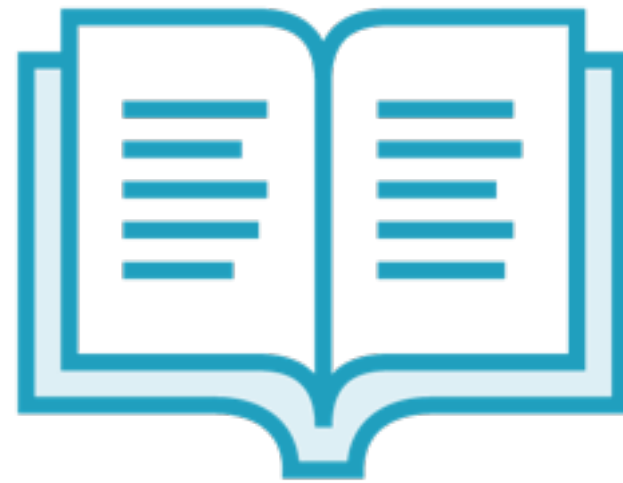
# Sentiment Lexicons

---

# Sentiment Lexicons Contain Word Metadata



Split text fragment  
into words



Look up each word in  
a **sentiment lexicon**



Arrive at polarity,  
intensity, mood,...

Sentiment lexicons form the core of virtually all  
sentiment analysis (rule-based and ML-based)

# Sentiment Lexicons Contain Word Metadata

## Dictionary

Lookup table for  
meanings of words

## Thesaurus

Lookup table for  
synonyms of words

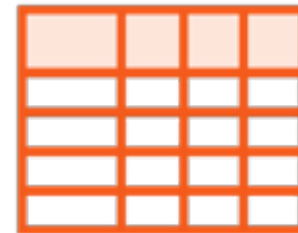
## Sentiment lexicon

Lookup table for  
intensity, polarity,...

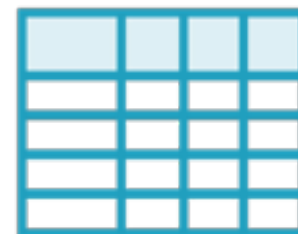
Like dictionaries, sentiment lexicons are extremely  
laborious to create

# Don't Try This at Home

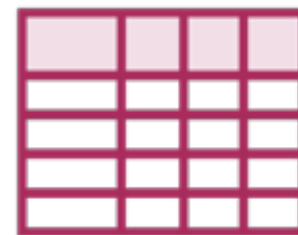
**Annotator #1**



**Annotator #2**



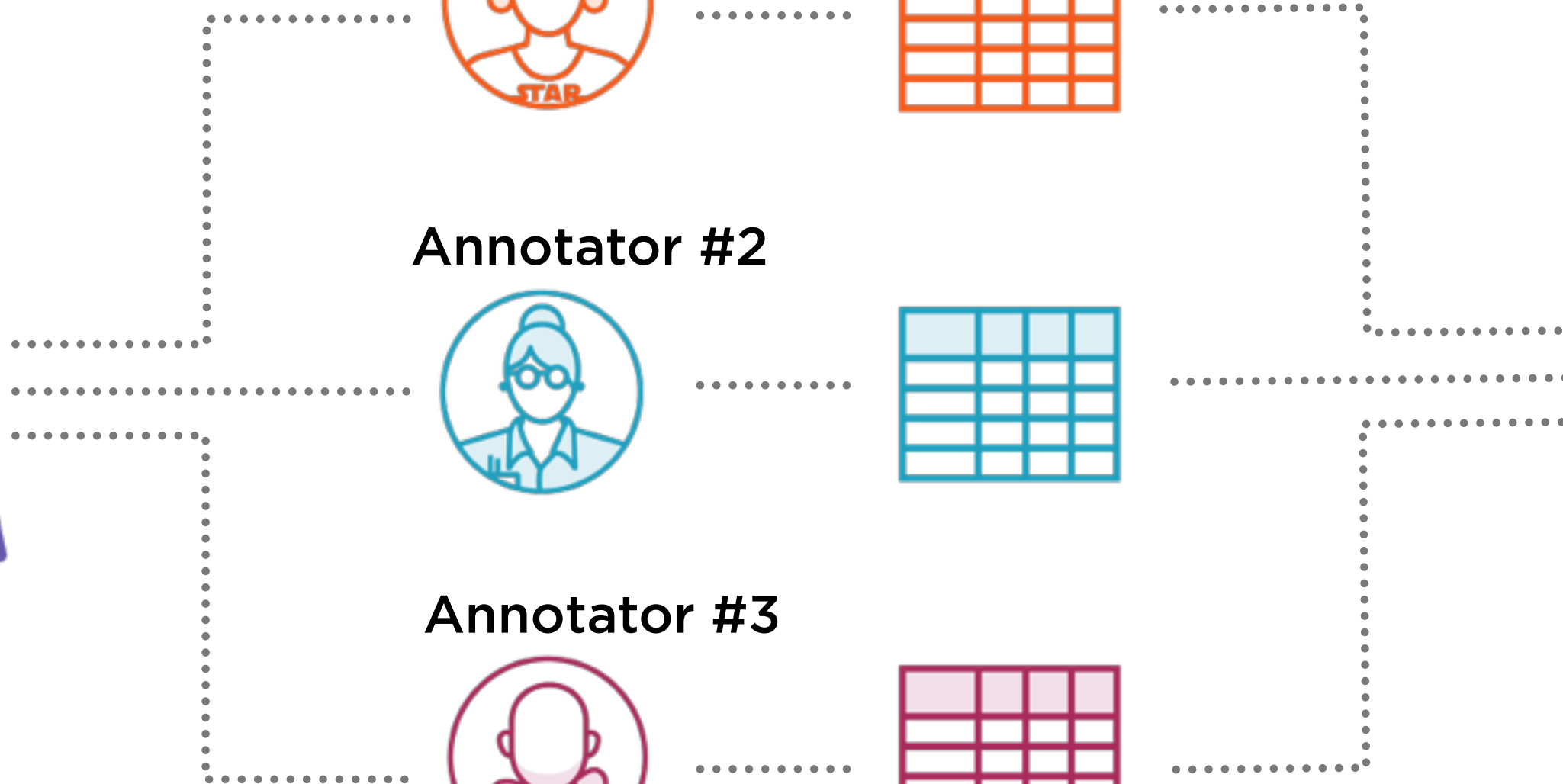
**Annotator #3**



**Word Sources**



**Sentiment  
Lexicon**





## Reliable, widely used sentiment lexicons

- Sentiwordnet
- MPQA
- LIWC
- General Inquirer

# Sentiment Lexicons Contain Word Metadata

## Polarity

Positive or negative?

## Subjectivity

Objective or subjective?

## Affective state

Emotions, moods, ...



# Affective State Enriches Rule Formulation

## Emotion

“Angry”, “Ecstatic”

## Mood

“Listless”, “Bored”

## Attitude

“Affectionate”, “Hostile”

## Personality trait

“Diligent”, “Methodical”

## Interpersonal stance

“Flirtatious”, “Warm”

# Building Is Hard, Using Is Easy



**Builder**

Building a sophisticated  
sentiment analysis system is **hard**



**User**

Using a sophisticated sentiment  
analysis system is **easy**

# Coming Up: Using VADER and Sentiwordnet



**VADER**

A sophisticated rule-based  
system



**Sentiwordnet**

A sophisticated sentiment  
lexicon

# Summary

**A naive rule-based classifier simply sums up individual word polarities**

**Rules can be added to deal with intensity and contrast**

**Sensitivity to context is a weak spot for rule-based systems**

**Sentiment lexicons are at the heart of it all**