Sentiment Analysis

Improving results with complex generated features

Lucas Silva & Dr.Pearl

Agenda

The field: Sentiment Analysis

Silva & Rocha: research

Evolution! Improving sentiment analysis

Methodology

Expected results

The field:

Sentiment analysis

"Sentiment analysis or opinion mining is the computational study of opinions, sentiments and emotions expressed in text"

(Liu, Bing)

Or...

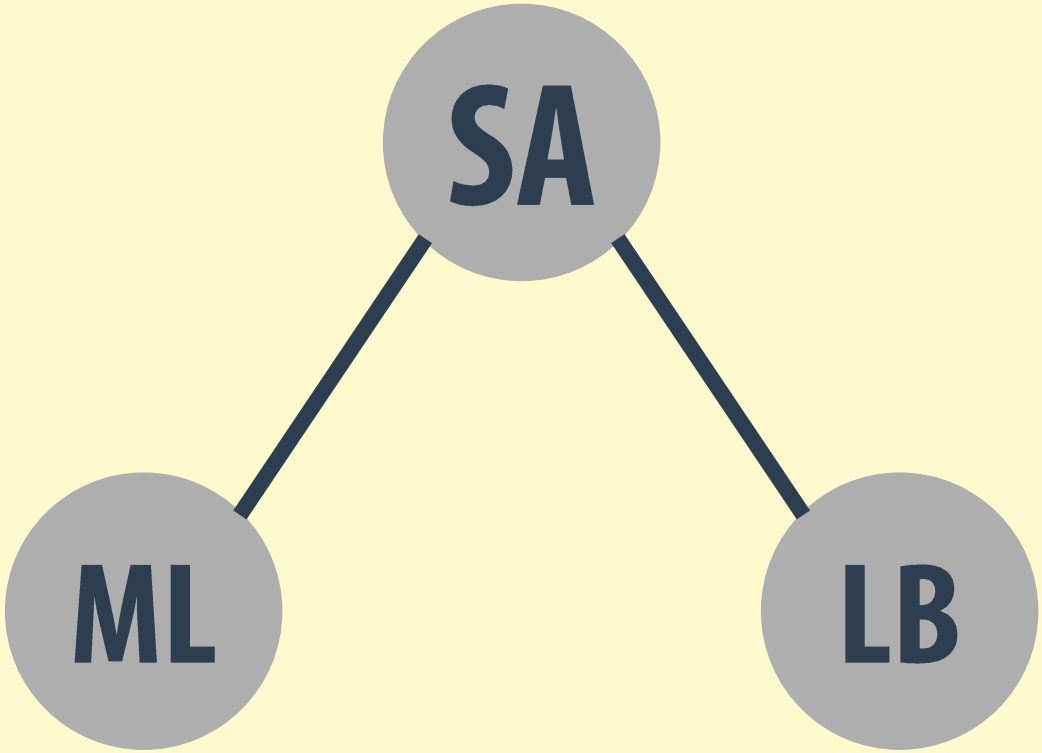
I loved the new Spider Man movie,
but I hated the poor performance of Andrew Garfield as Peter Parker

I loved the new Spider Man movie, but I hated the poor performance of Andrew Garfield as Peter Parker

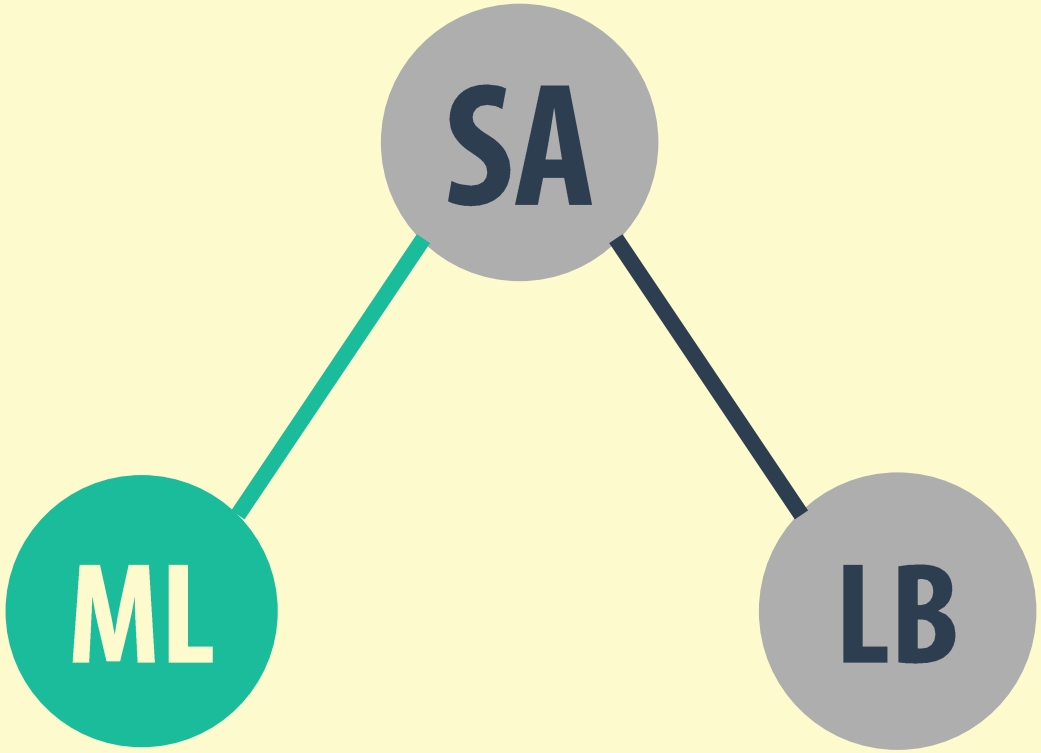
I loved the new Spider Man movie,

but I hated the poor performance of Andrew Garfield as Peter Parker

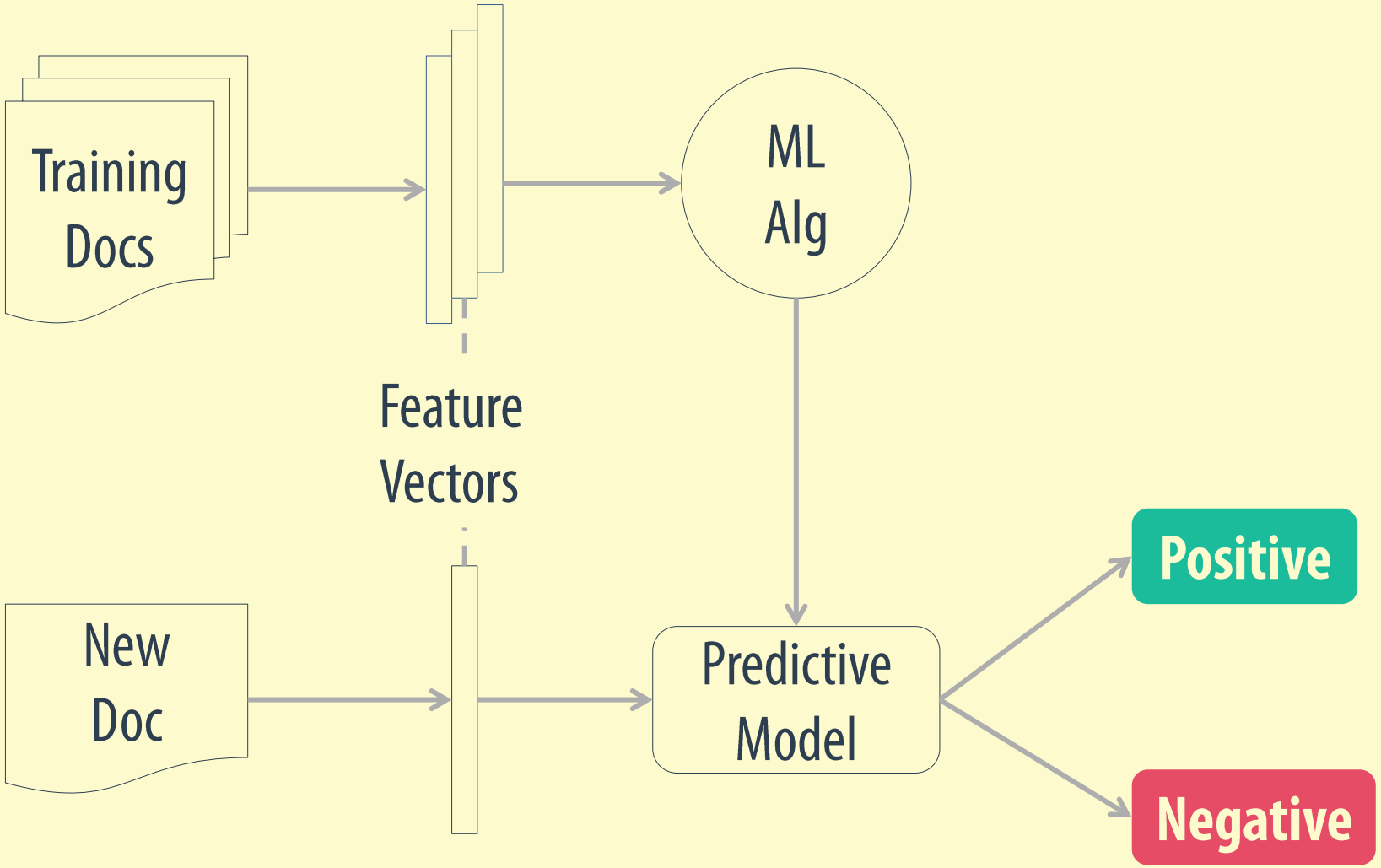
Organizing the field



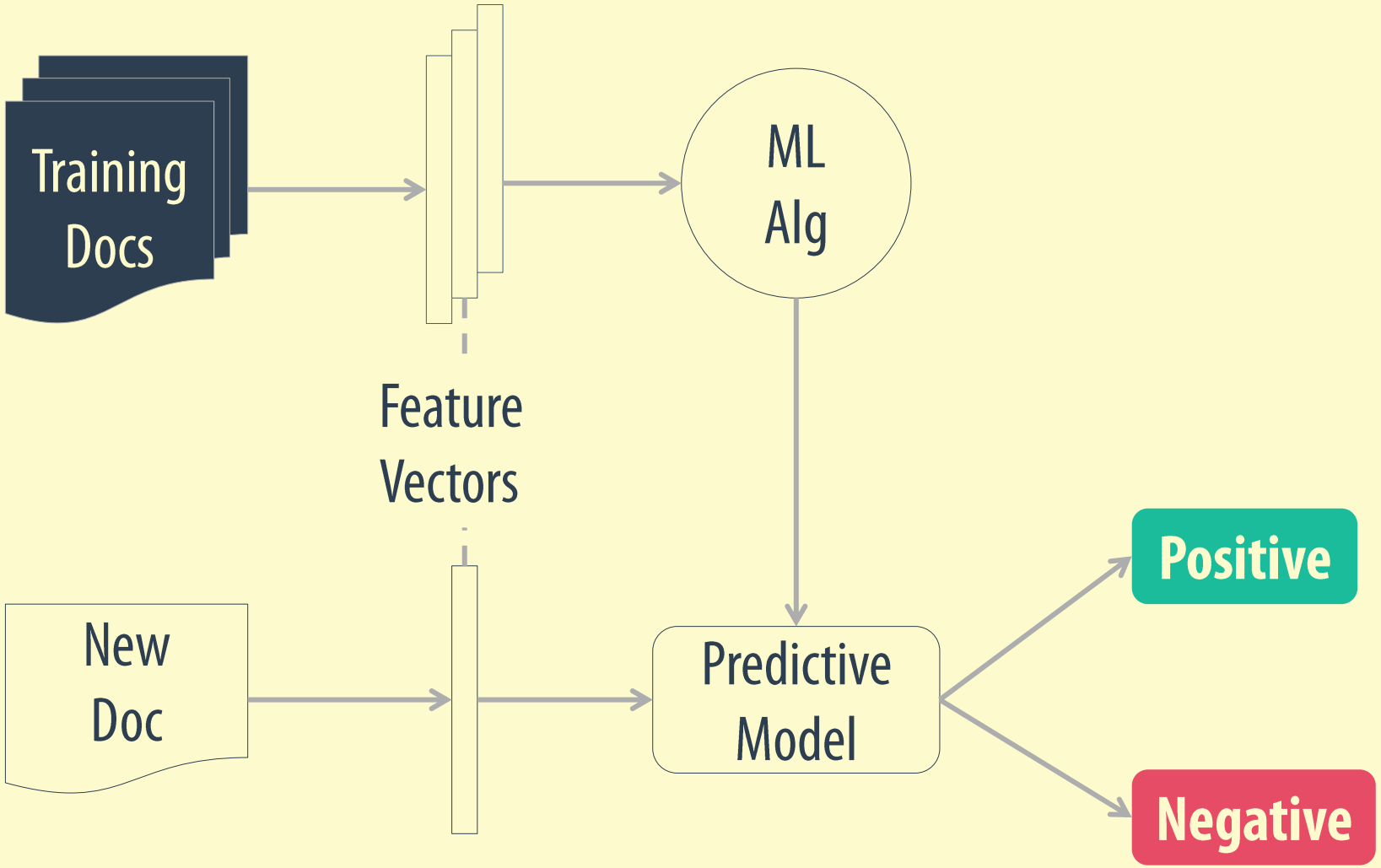
Organizing the field



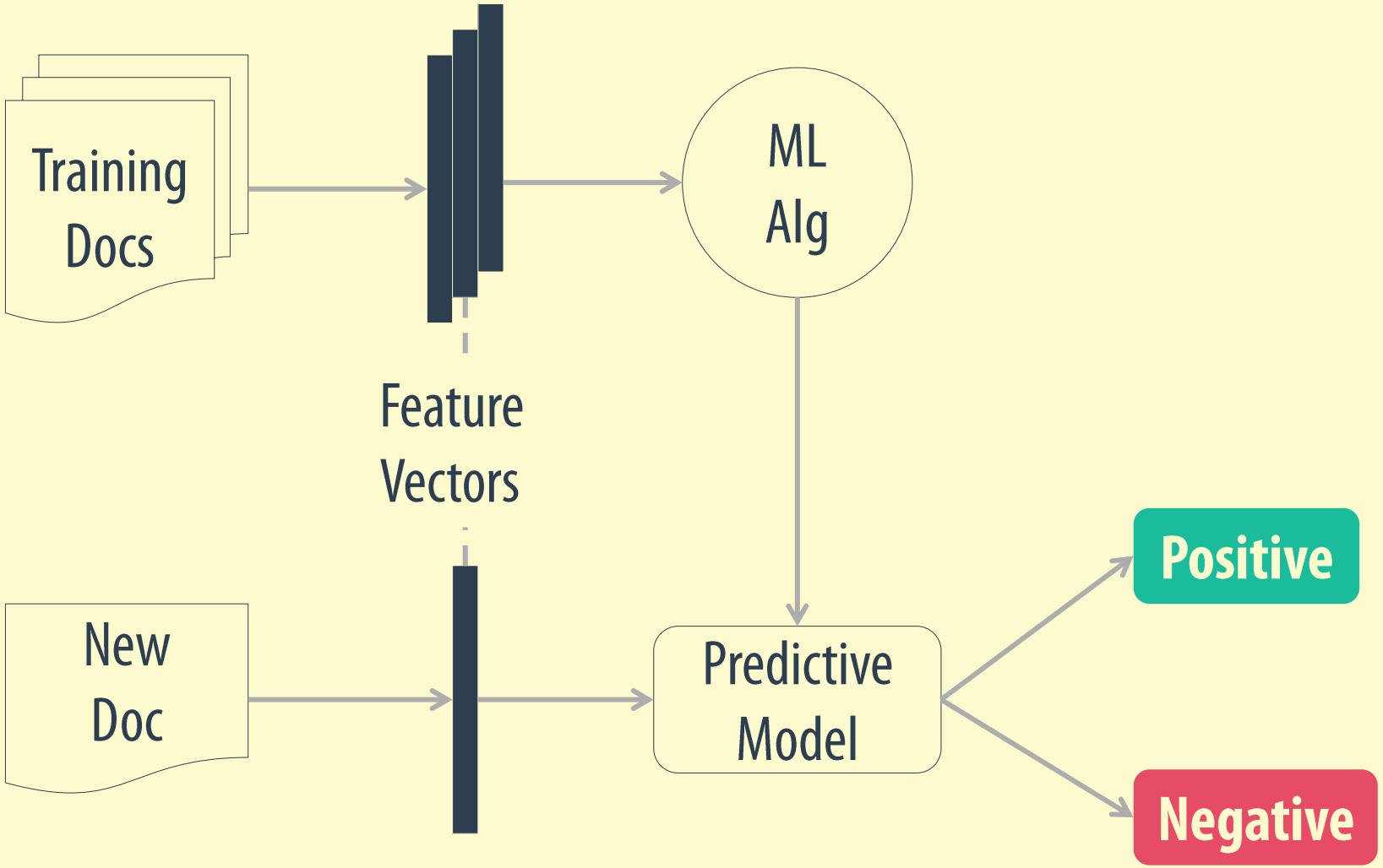
Machine Learning

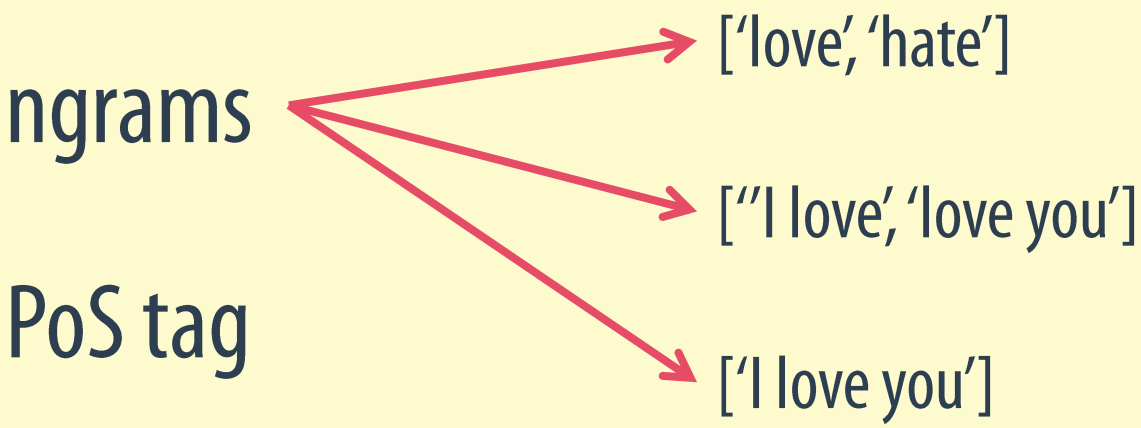


Machine Learning



Machine Learning





PoS tag

Prior polarity

Polarity shifters

Specific sets of words

ngrams

PoS tag → ['I_PRP', 'love_VBP', 'you_PRP']

Prior polarity

Polarity shifters

Specific sets of words

ngrams

PoS tag

Prior polarity → ['love_pos']

Polarity shifters

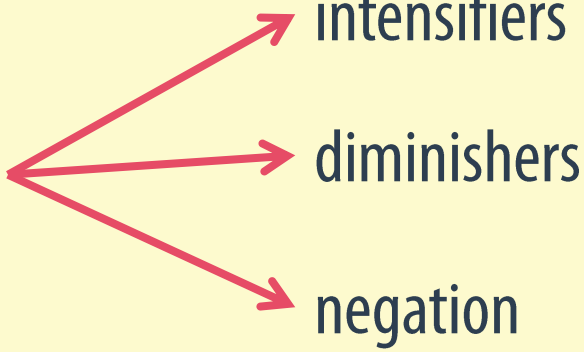
Specific sets of words

ngrams

PoS tag

Prior polarity

Polarity shifters



Specific sets of words

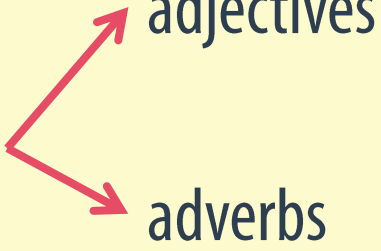
ngrams

PoS tag

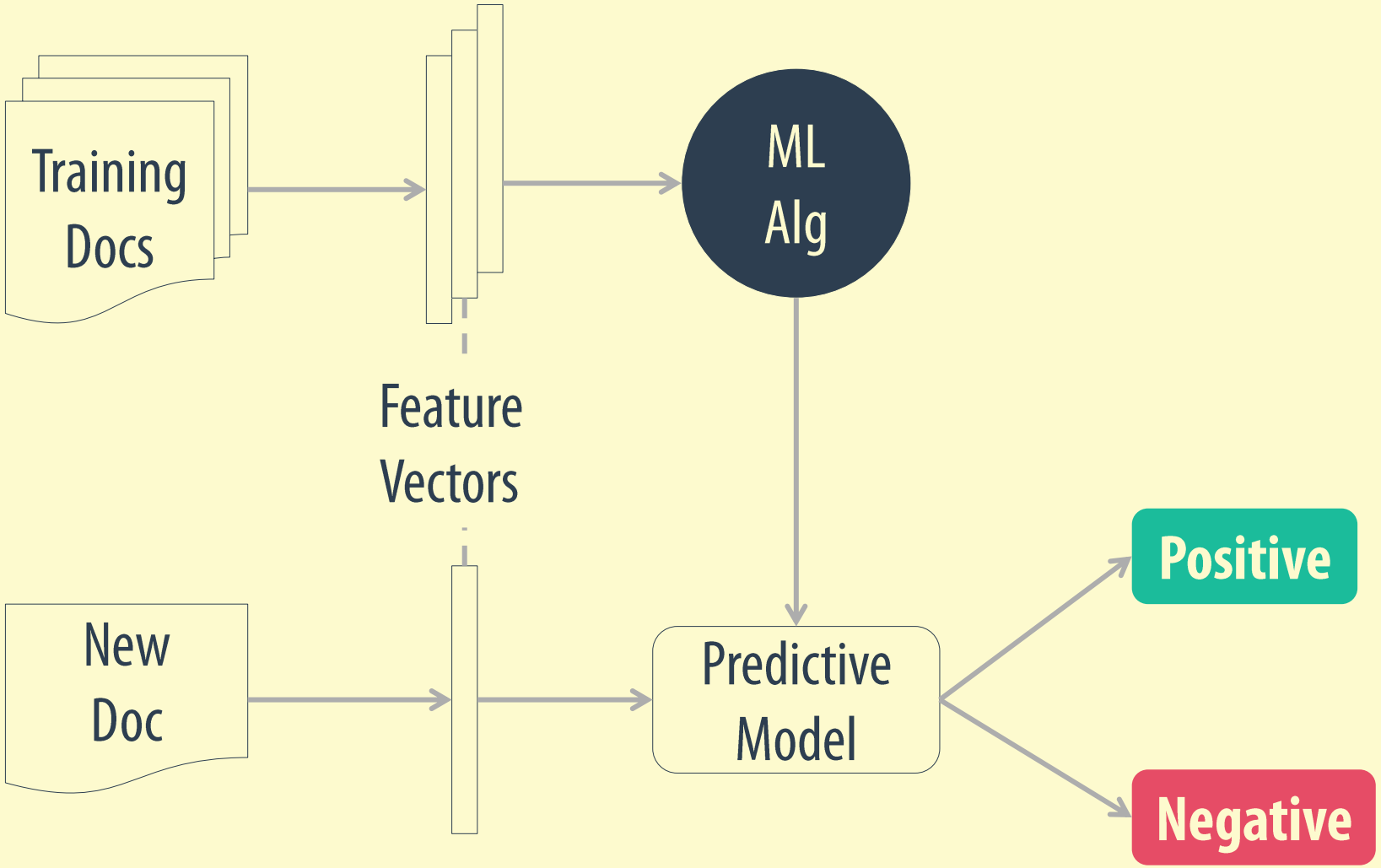
Prior polarity

Polarity shifters

Specific sets of words



Machine Learning



Main algorithms are...

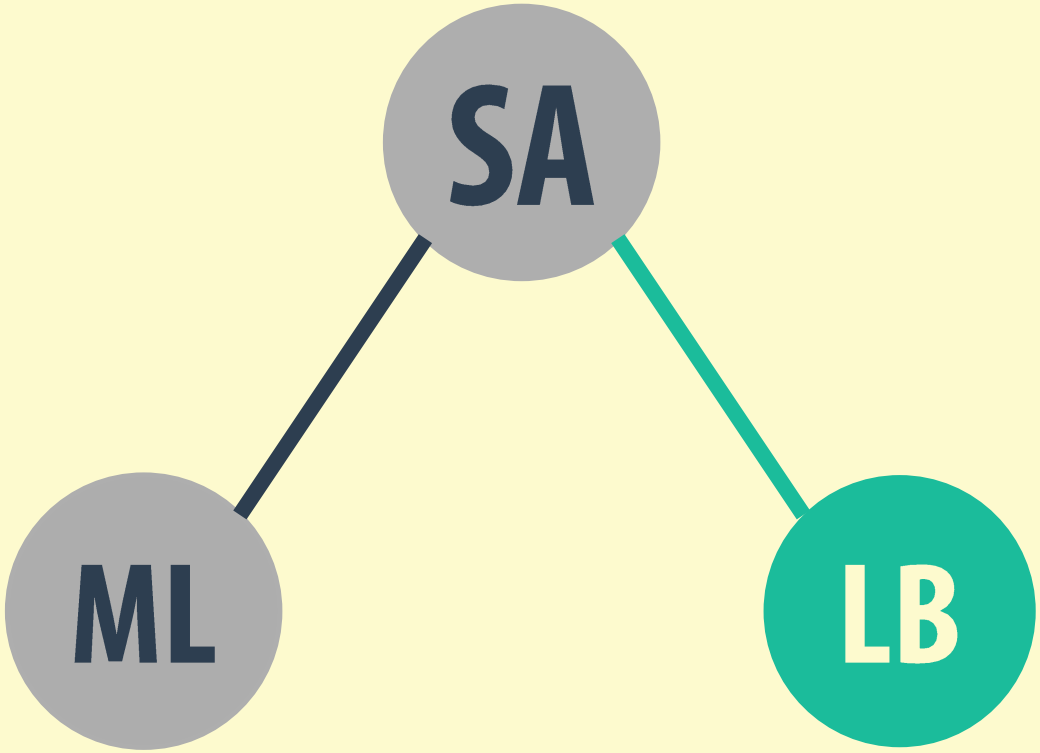
Naive Bayes

Assumes that all the features are independent and equally important

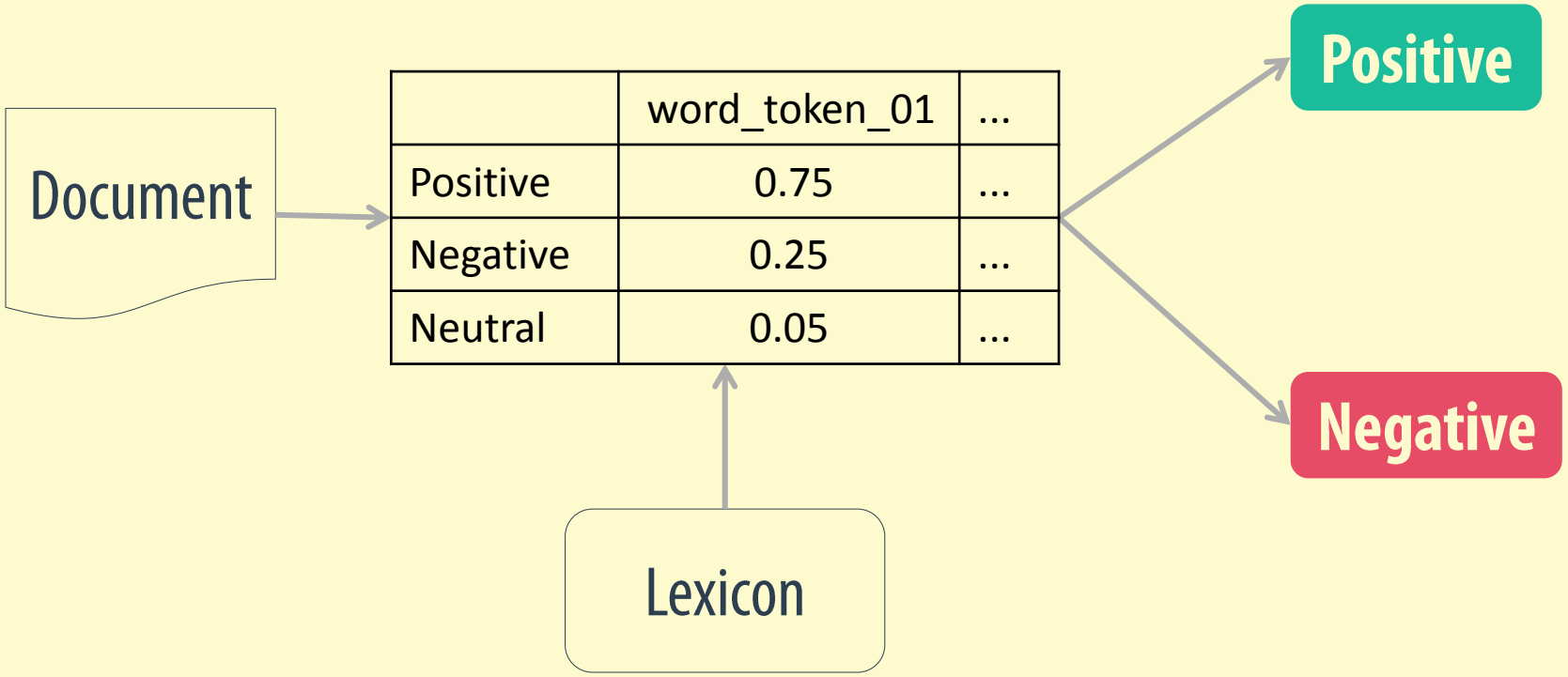
Support Vector Machines

Has the ability to define the most important features

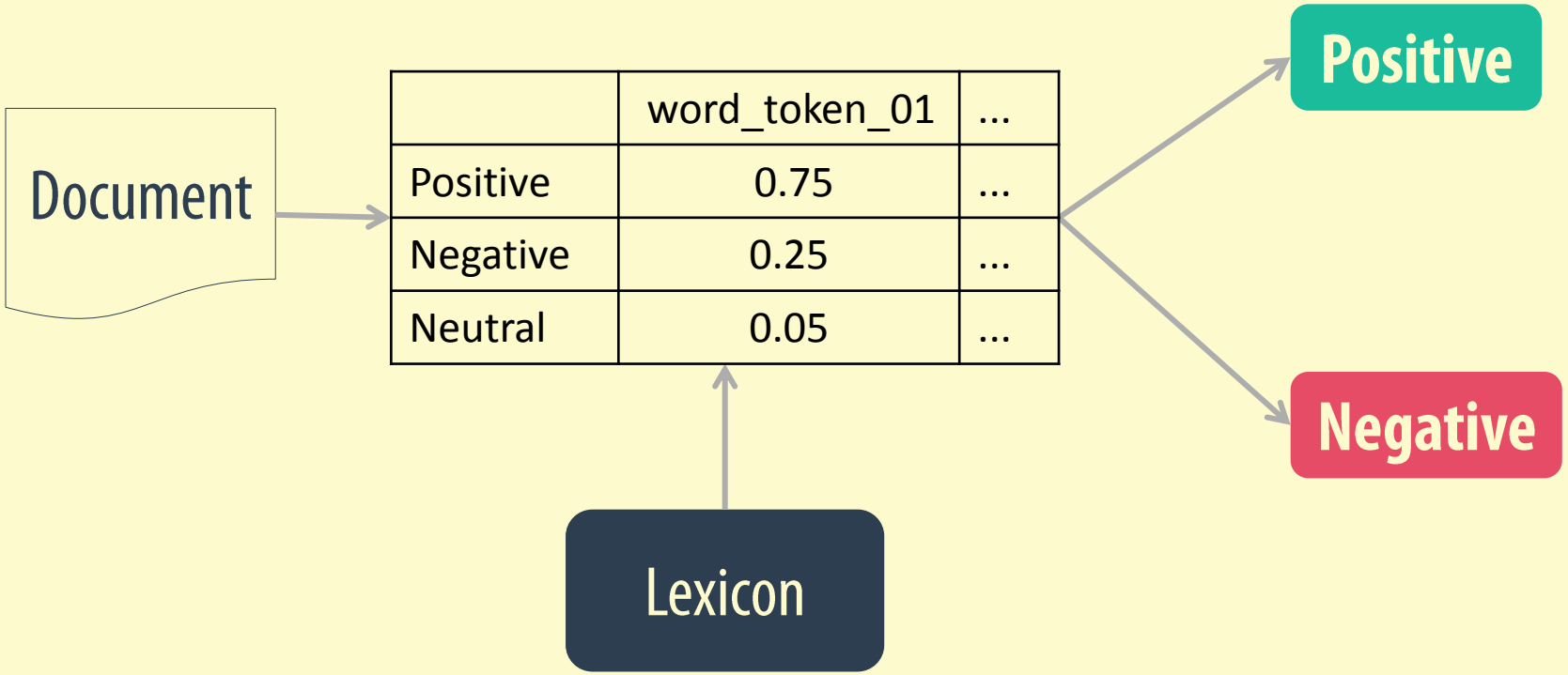
Organizing the field



Lexicon Based



Lexicon Based



Main lexicons are...

General Inquirer

WordNet

SentiWordNet

Valence Shifters

Modifiers: intensifiers and diminishers

This is really good!

Valence Shifters

Modifiers: intensifiers and diminishers

This is really **good**!

+ 0.78

Valence Shifters

Modifiers: intensifiers and diminishers

Valence Shifters

Negation: switch

This is not good!

Valence Shifters

Negation: switch

This is not **good**!

+ 0.78

Valence Shifters

Negation: switch

Valence Shifters

Negation: shift

Valence Shifters

window of words x window of structure

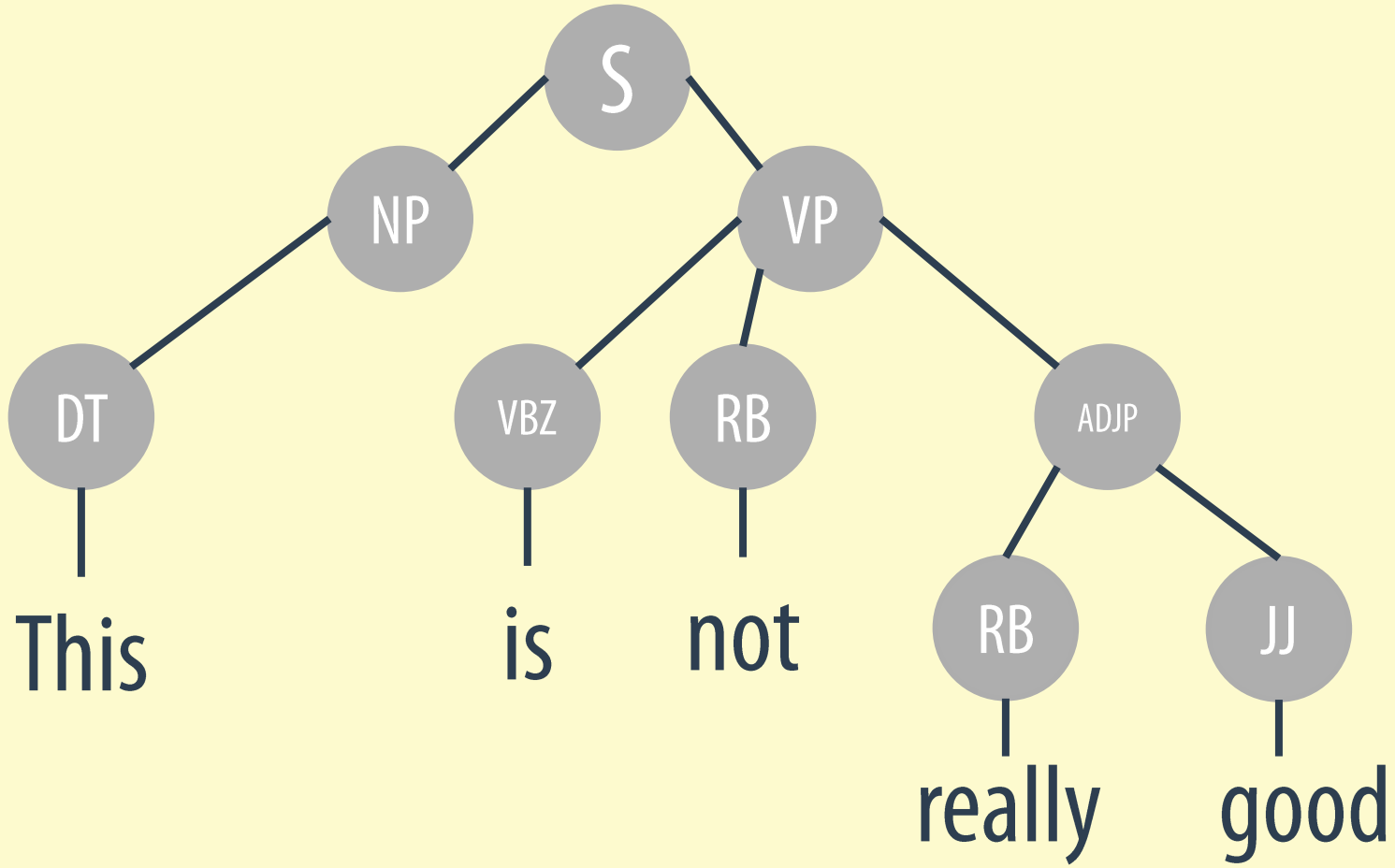
This is not really good!

Valence Shifters

Window: words **X** structure

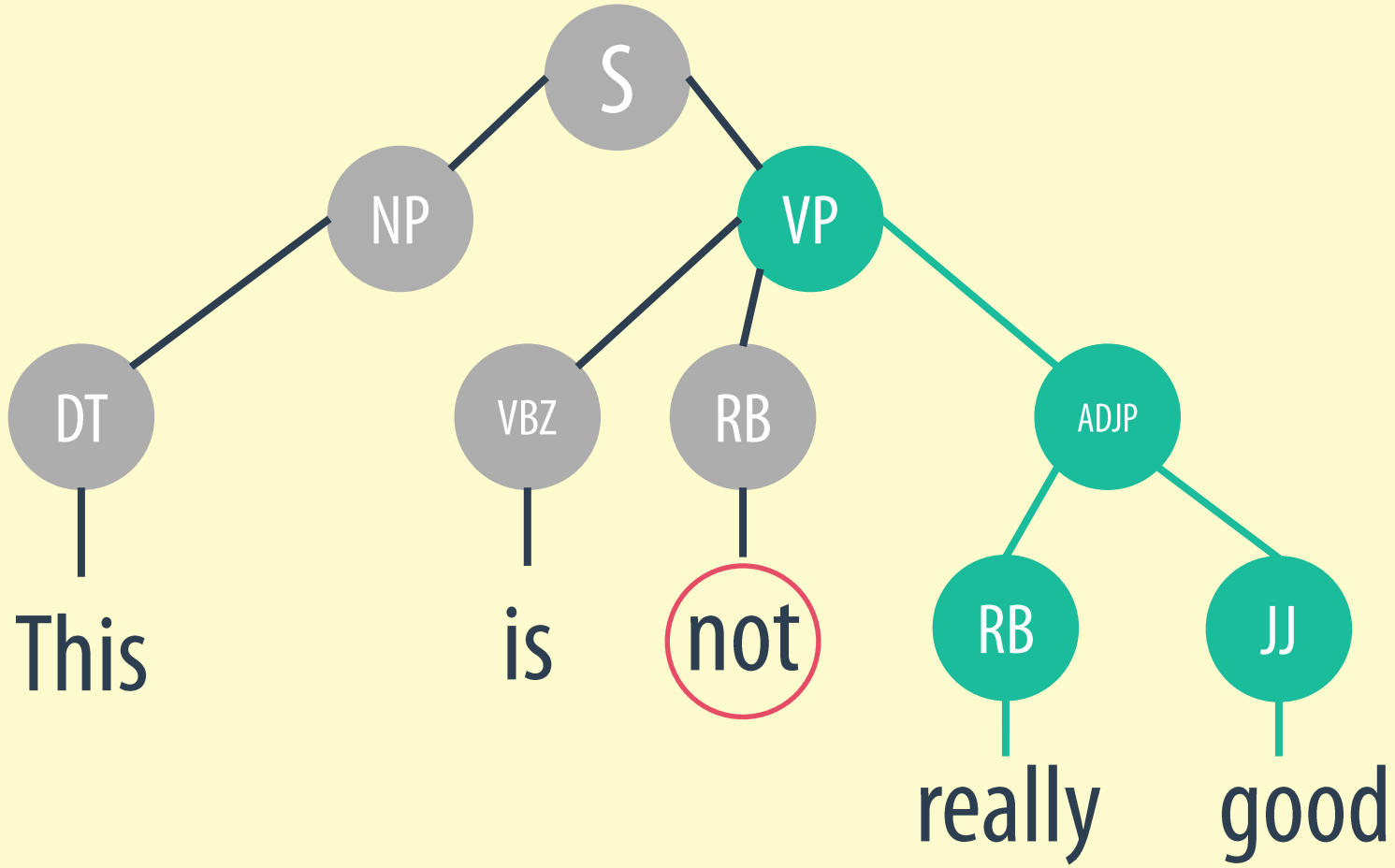
Valence Shifters

Window: words **X** structure



Valence Shifters

Window: words **X** structure



Silva&Rocha:

research

Dataset

Topic: entrepreneurship/startups

Automatically acquired using crawler

180 articles with 4563 sentences

Goal

Show that it's possible to create accurate SA classifiers using data provided by untrained annotators

Pipelines

- Initial data classification
- Annotation and final classification

Methodology

Pipelines

- Initial data classification
- Annotation and final classification

Classifier: naive Bayes

Features: unigrams

Approach: Generated different training data and compare results of the classifiers trained with them

Evolution!

Improving sentiment analysis

Goal

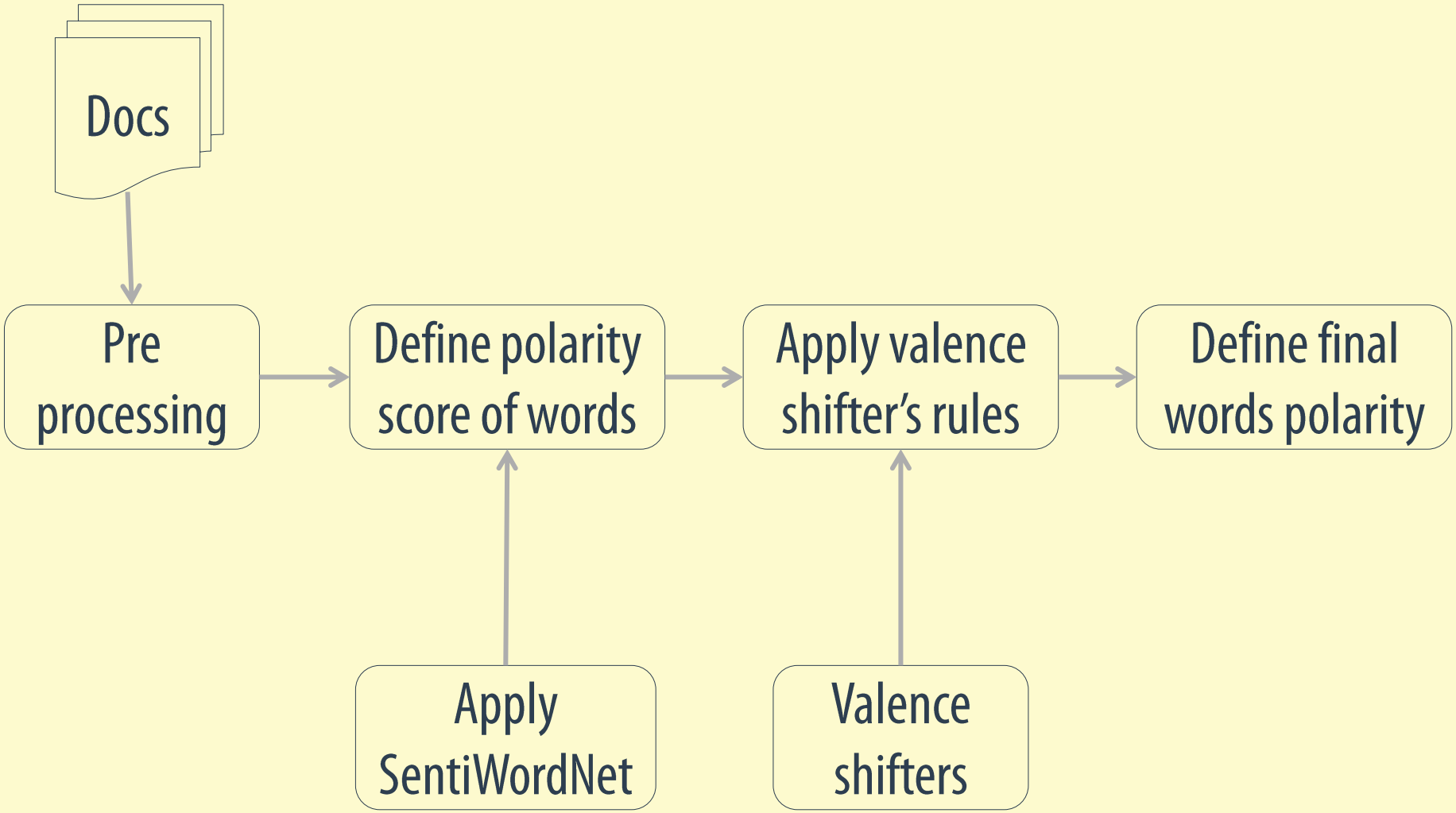
To study how the process to generate word prior polarity feature impacts the accuracy of the sentiment classification task

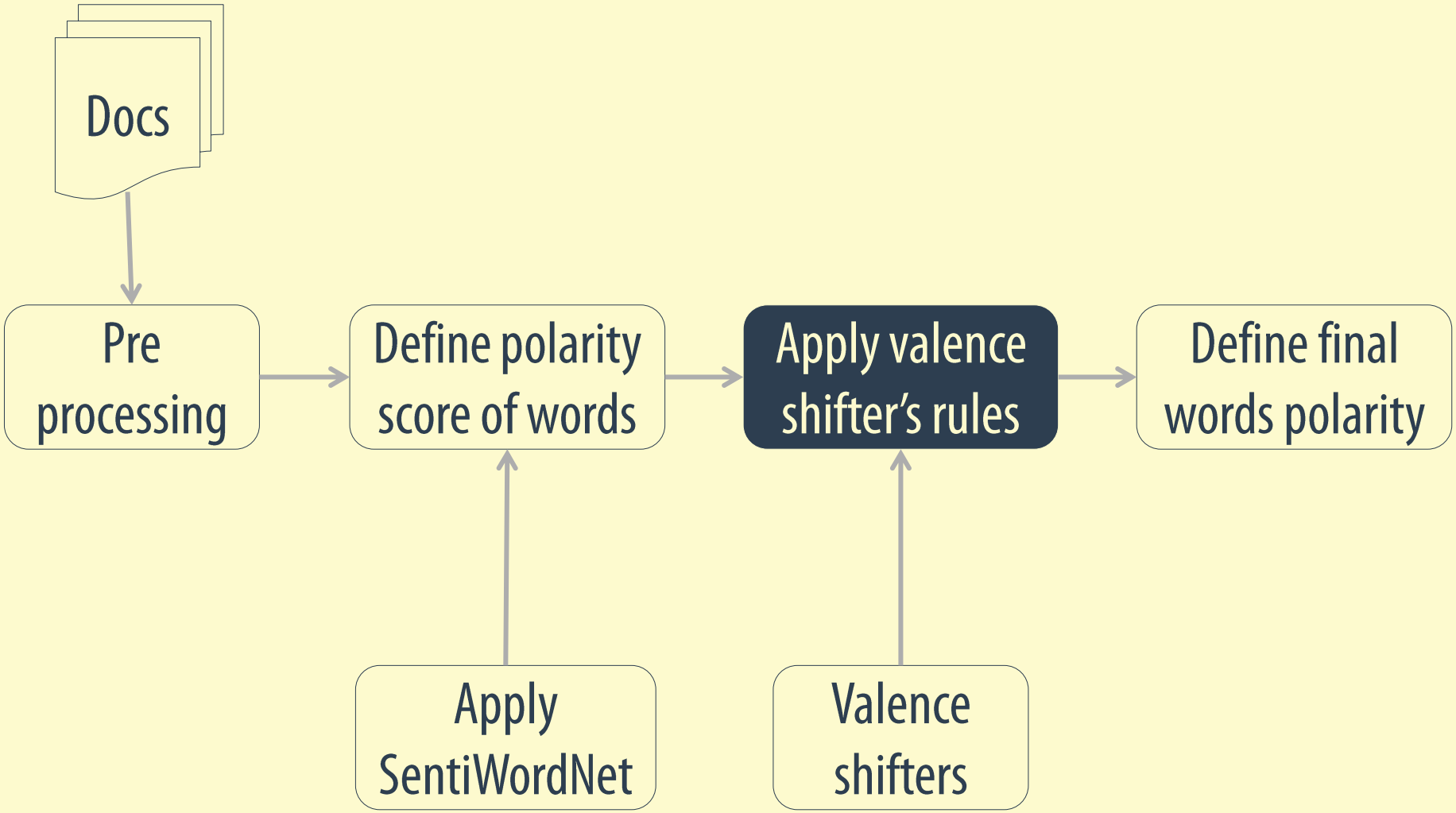
Methodology

1. Framework to generate complex features
2. Sentiment classification task
3. Experiment

Methodology

1. **Framework to define polarity of words**
2. Sentiment classification task
3. Experiment


Docs
Pre processing
Define polarity score of words
Apply valence shifter's rules
Define final words polarity
Apply SentiWordNet
Valence shifters


Docs
Pre processing
Define polarity score of words
Apply valence shifter's rules
Define final words polarity
Apply SentiWordNet
Valence shifters

This is **good**	.75
This is **really good**	.1 + .75
This is **not good**	1 - |.75|
This is **not really good**	1 – |.1 + .75|
This is **bad**	-.625

Methodology

1

Framework to define polarity of words

2

Sentiment classification task

3

Experiment

Classification

Features

this	is	good	really	not	bad
0.0	0.0	.75	0.0	0.0	0.0
0.0	0.0	.85	0.0	0.0	0.0
0.0	0.0	-.25	0.0	0.0	0.0
0.0	0.0	-.15	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	-.625

Classification

Classifier: Support Vector Machines

10-fold cross validation to define best params

Methodology

1

Framework to define polarity of words

2

Sentiment classification task

3

Experiment

Experiment

Compare:

Baseline (Silva&Rocha)

Support Vector Machines with different windows for shifters

Precision, recall and F-score to define accuracy

Expected results

We believe that...

Negation shift >> Negation switch

Window of structure >> Window of words

SVM >> naive Bayes

Thank you!