MASTER'S DEGREE THESIS

Sentiment Analysis For Touristic Attractions: A Case Study On The Alhambra

Ana Valdivia

19 September 2016

Advisor: Salvador García

Department of Computer Science and Artificial Intelligence

Universidad de Granada

CONTENTS

1. INTRODUCTION

BACKGROUND AND MOTIVATION

1. INTRODUCTION

Tourism

Source: http://fr.123rf.com/photo_7306498_avion-de-tourisme-icones-sentier-volant-dans-le-monde-entier.html

1. INTRODUCTION

Web 2.0

Source: https://www.emaze.com/@AZILZITO/WEB-2.0 l

1. INTRODUCTION

Data

Source: https://www.youtube.com/watch?v=wWcgYZWCAXg

1. INTRODUCTION

What tourists think about touristic attractions?

1. INTRODUCTION

OBJECTIVES

1. INTRODUCTION

OBJECTIVES

First approach to sentiment analysis into touristic attractions domain

1. INTRODUCTION

More precisely…

1. Download and **analyse** reviews

More precisely…

1. **Download** and **analyse** reviews

2. **Study** correlation between **human** and **machine sentiment**

1. INTRODUCTION

More precisely…

1. **Download** and **analyse** reviews

2. **Study** correlation between **human** and **machine sentiment**

3. **Classify** opinions

1. INTRODUCTION

More precisely…

1. **Download** and **analyse** reviews
2. **Study** correlation between **human** and **machine sentiment**
3. **Classify** opinions
4. **Dicover** interesting **patterns** in **negative** opinions

2. DATA

Where do we get the data?

2. DATA

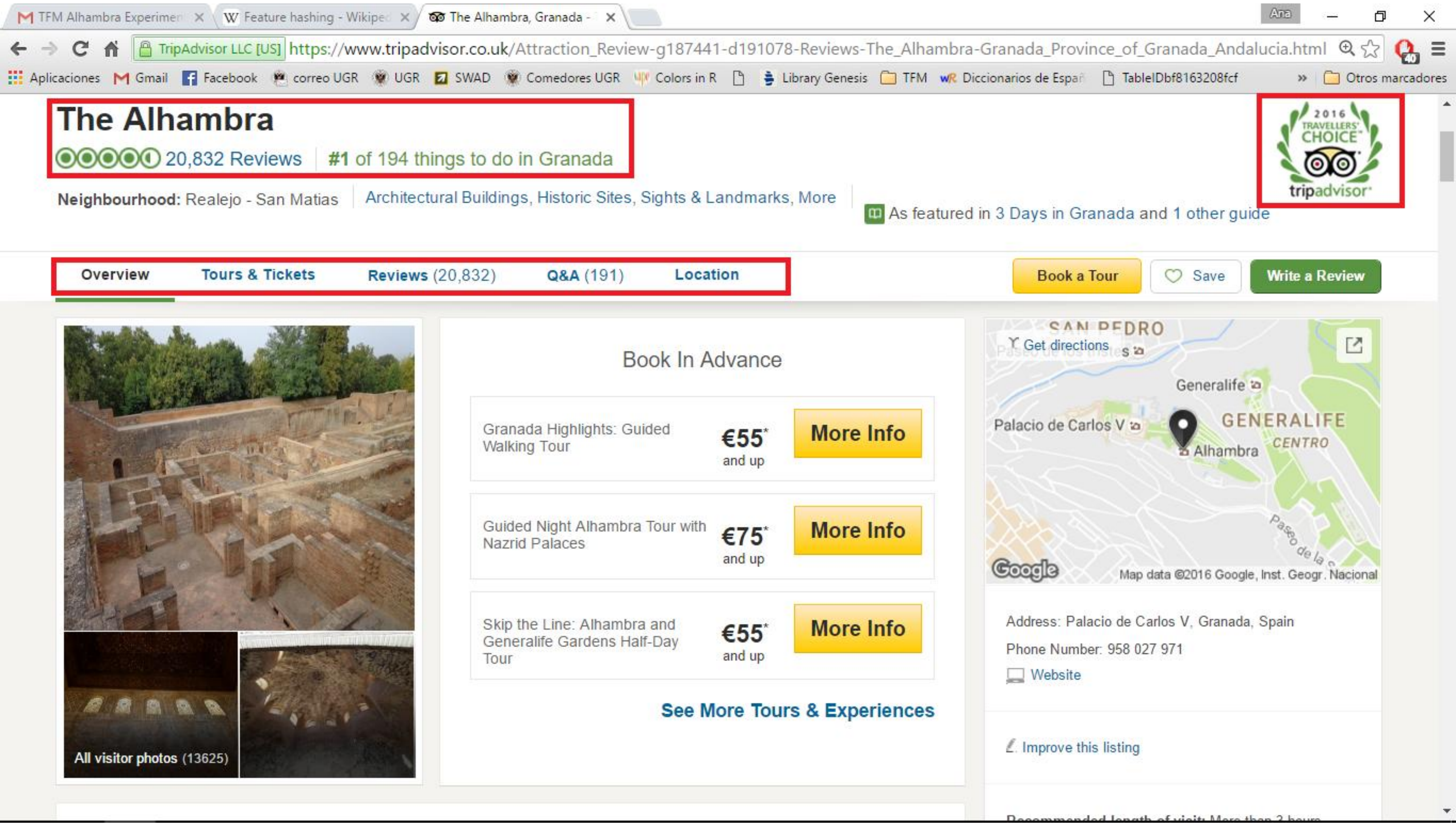
TripAdvisor logo

2. DATA

The Alhambra

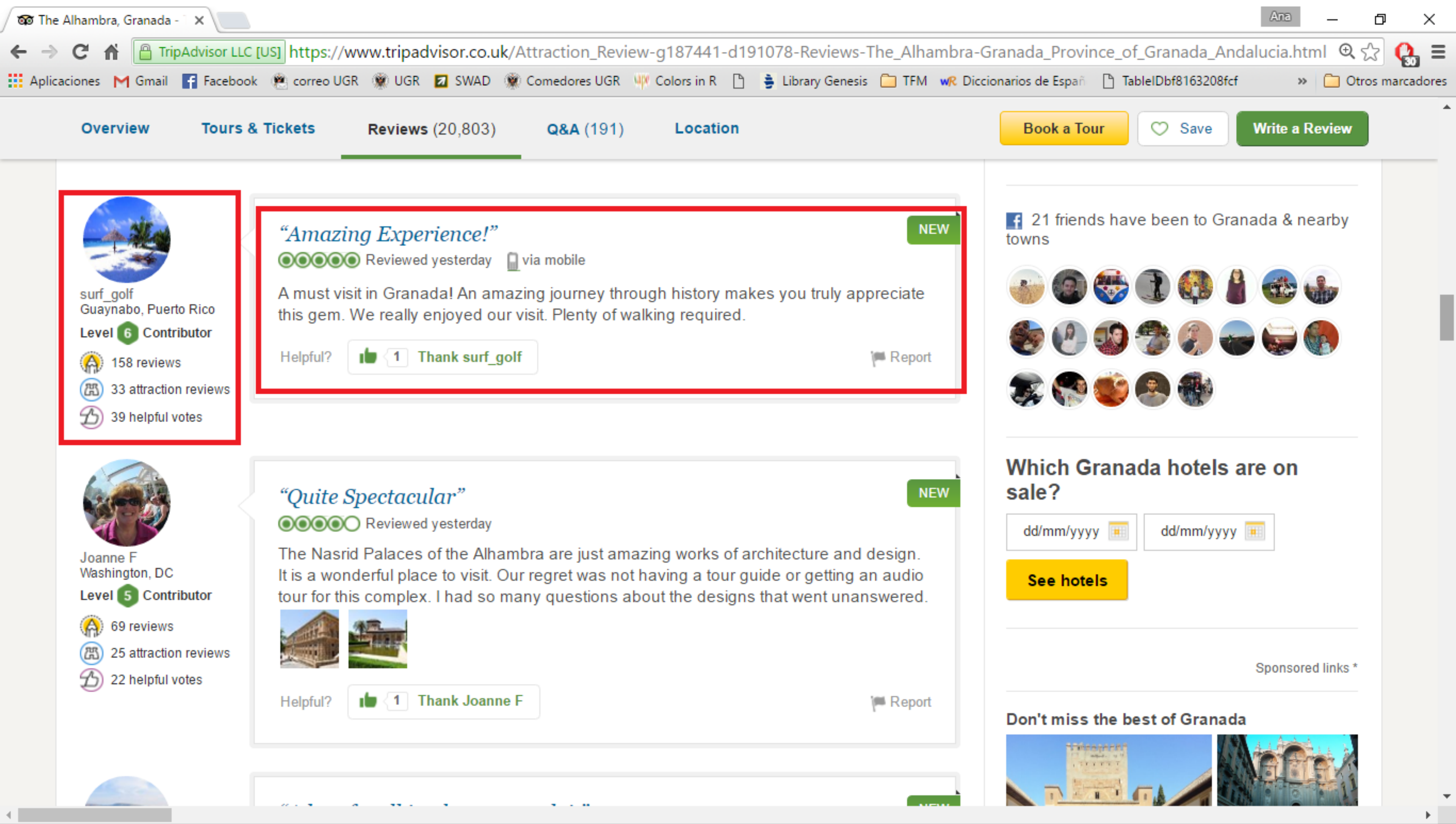
2. DATA

SCRAPPING



2. DATA

SCRAPPING



2. DATA

8140 instances

TripAdvisorAlhambra data set

2. DATA

STRUCTURE

8140 instances

10 features

```
id
username
location
userop
quote
review
quote+review
rating
date
page
```

TripAdvisorAlhambra data set

2. DATA

STRUCTURE

8140 instances

TripAdvisorAlhambra data set

10 features

```
id
username
location
userop
quote
review
quote+review
rating
date
page
```

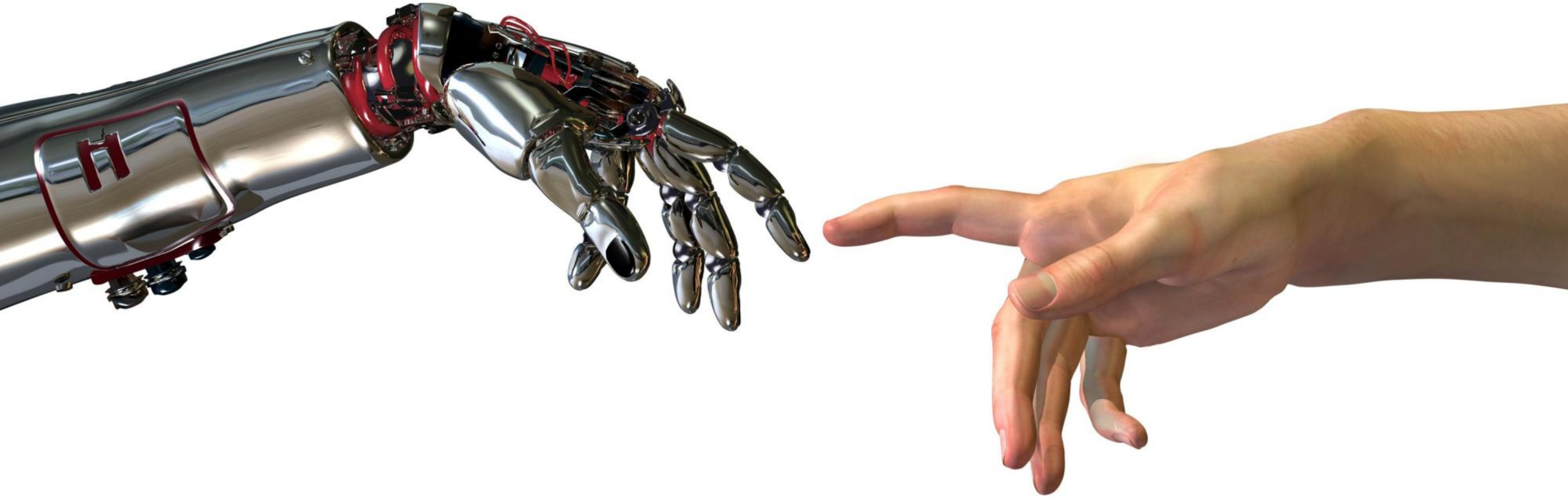
2 class labels

2. DATA

Class labels

SentimentCoreNLP

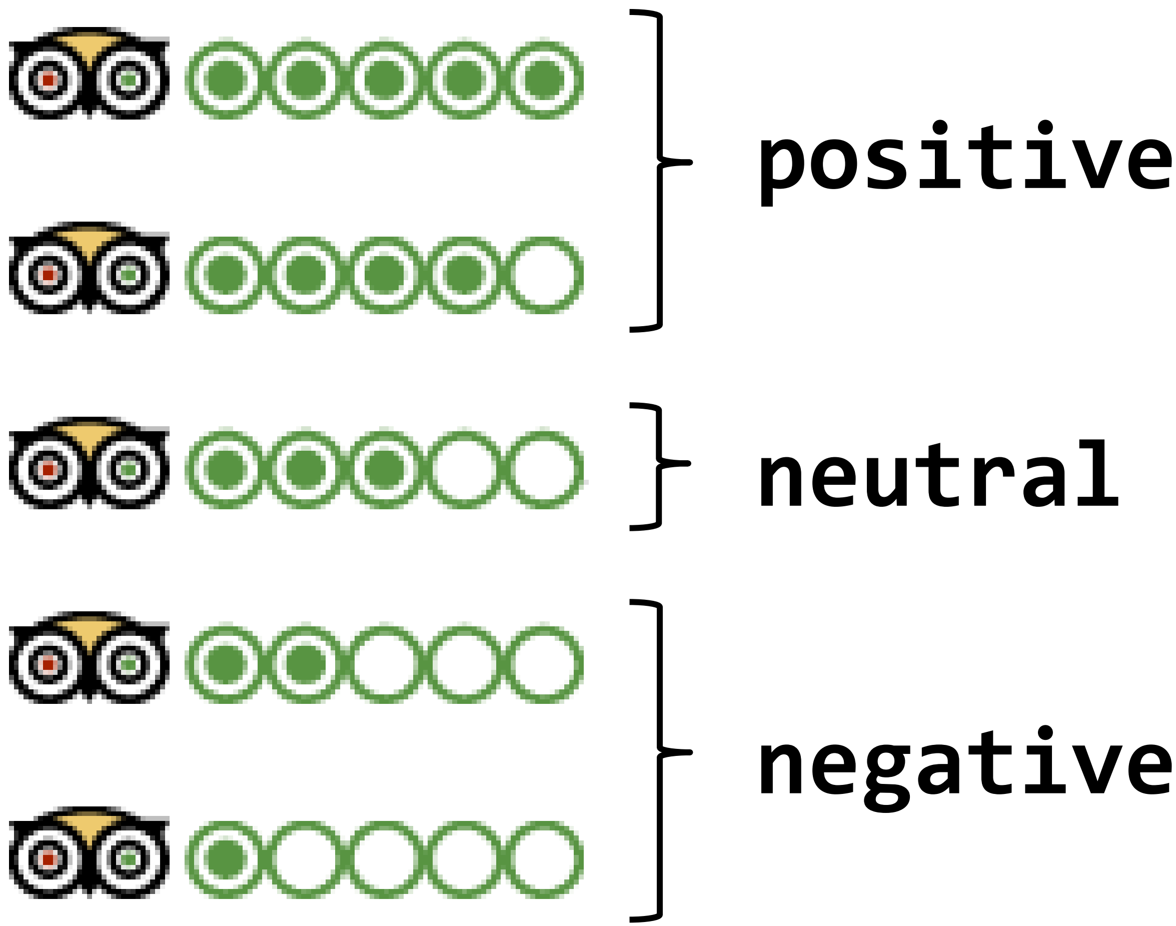
SentimentValue



2. DATA

STRUCTURE

SentimentValue



2. DATA

STRUCTURE

reviews

SentimentCoreNLP

2. DATA

STRUCTURE

reviews

SentimentCoreNLP

CoreNLP toolkit

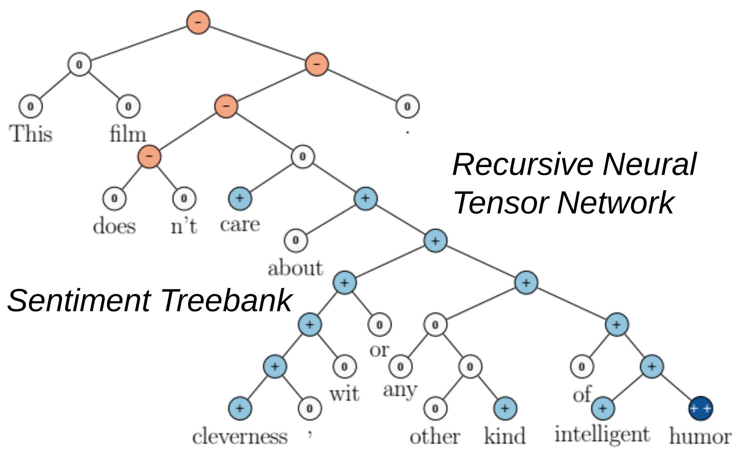
2. DATA

STRUCTURE

reviews

SentimentCoreNLP

CoreNLP toolkit



Recursive Neural Tensor Network

Stanford Sentiment Treebank

2. DATA

STRUCTURE

reviews

"Very Moorish ;-)"

Reviewed 30 June 2016

Judging by the queues pre booking a must (which we did)
Get in the 'timed queue' for the Palaces early and take a cold beer that can be bought from the round cabin near the entrance to the Armoury (next door)
Beautiful gardens to walk around (& a lot of walking is involved!) and very intricate plasta work adorn the many walls and chambers. A guide book (or guide) will help as very little 'public' information is proffered as you wander around the site. Possibly you could rush around it, but I would suggest needing a minimum of 4 hours. Car park quite expensive, it charges by minutes though to amaximum of €18 so no rush to get out early.

Visited June 2016

Less

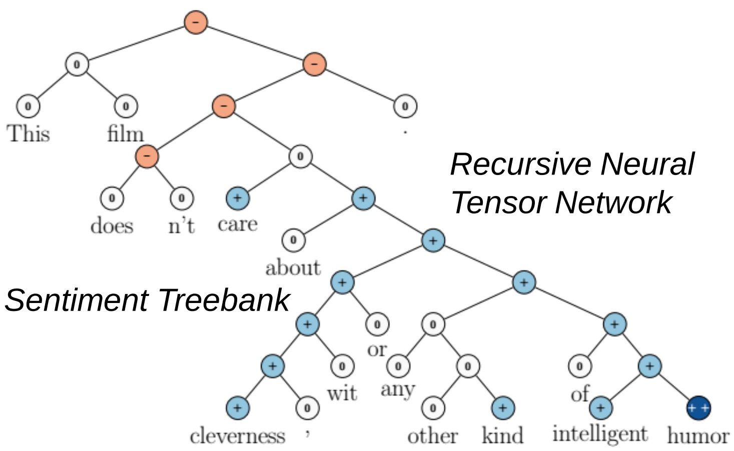
Helpful? 1 Thank Lindonian Report

Ask Lindonian about The Alhambra

This review is the subjective opinion of a TripAdvisor member and not of TripAdvisor LLC

SentimentCoreNLP

CoreNLP toolkit



Recursive Neural Tensor Network

Stanford Sentiment Treebank

Σ

2. DATA

STRUCTURE

SentimentCoreNLP

reviews

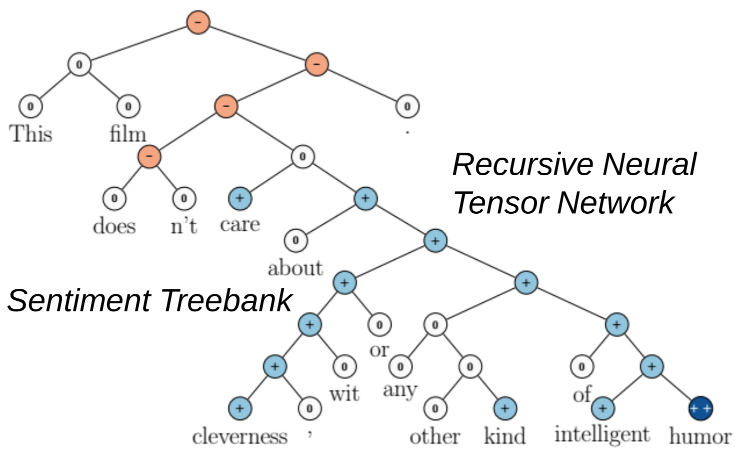
positive

neutral

negative

Σ

CoreNLP toolkit



Recursive Neural Tensor Network

Stanford Sentiment Treebank

ANALYSIS OF CORRELATION BETWEEN CLASS LABELS

3. CLASSIFICATION ANALYSIS

CORRELATION CLASS LABELS

	SentimentCoreNLP			
SentimentValue	positive	neutral	negative	Total
positive	4,049	1,071	2,508	7,628
neutral	51	32	260	343
negative	5	6	158	169
Total	4,105	1,109	2,926	8,140

Table 4.2: Correlation between SentimentValue and SentimentCoreNLP

53.08 % of coincidence

3. CLASSIFICATION ANALYSIS

CORRELATION CLASS LABELS

	SentimentCoreNLP			
SentimentValue	positive	neutral	negative	Total
positive	4,049	1,071	2,508	7,628
neutral	51	32	260	343
negative	5	6	158	169
Total	4,105	1,109	2,926	8,140

Table 4.2: Correlation between SentimentValue and SentimentCoreNLP

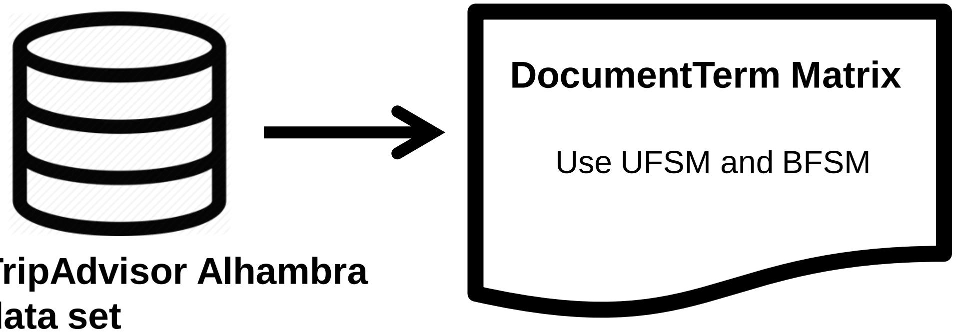
93.49 % of coincidence

CLASSIFICATION ANALYSIS

3. CLASSIFICATION ANALYSIS

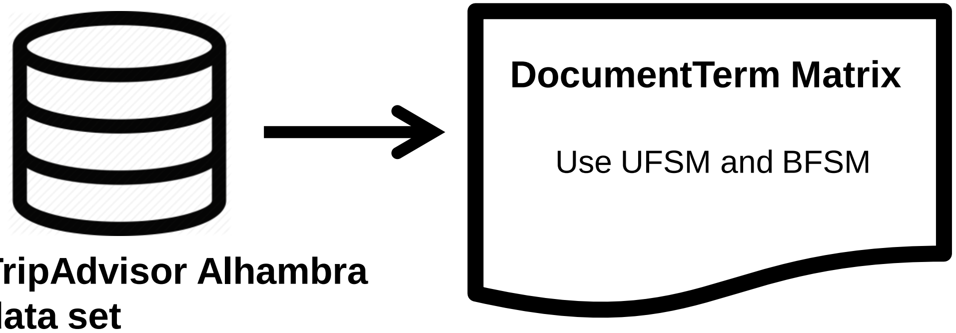
TripAdvisor Alhambra data set

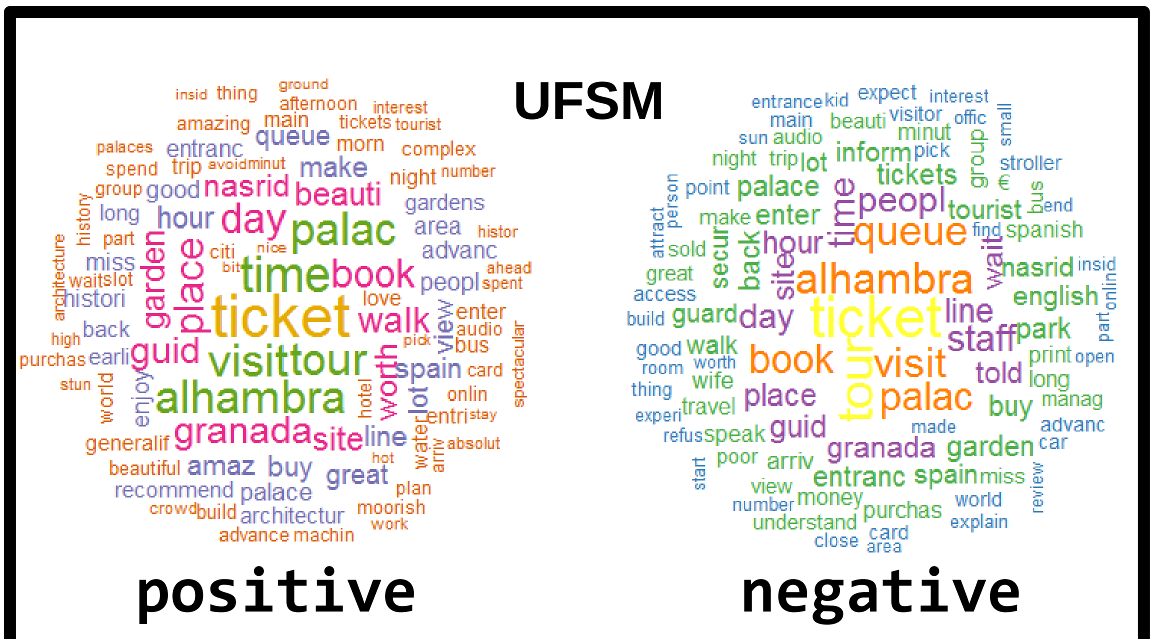
3. CLASSIFICATION ANALYSIS



TripAdvisor Alhambra data set

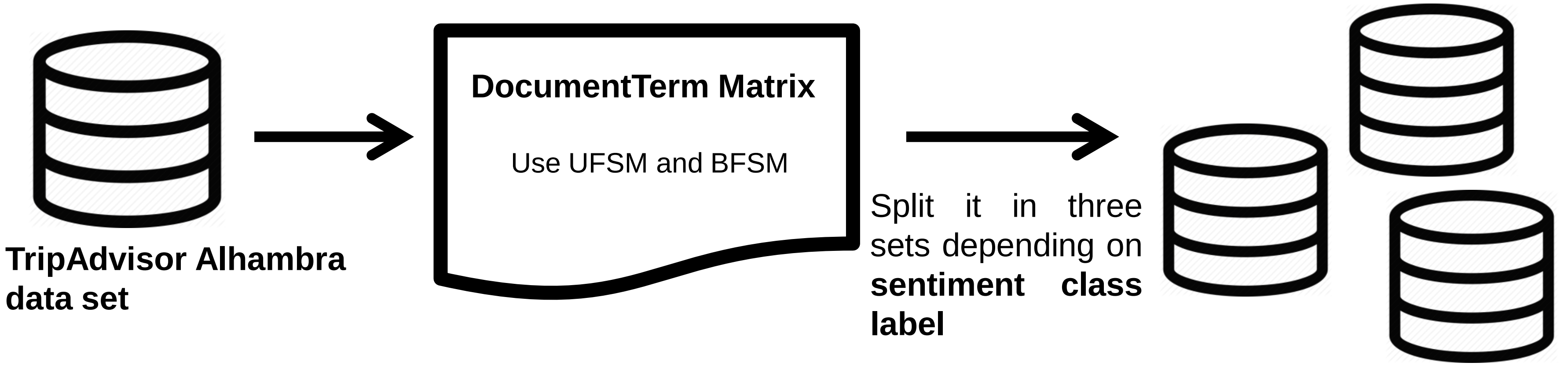
3. CLASSIFICATION ANALYSIS



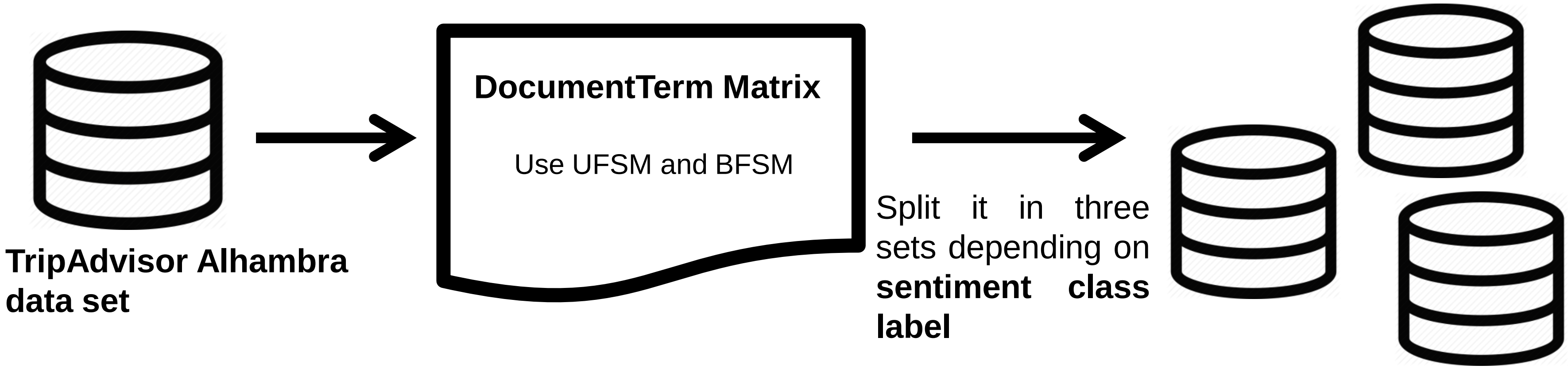


3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS



3. CLASSIFICATION ANALYSIS



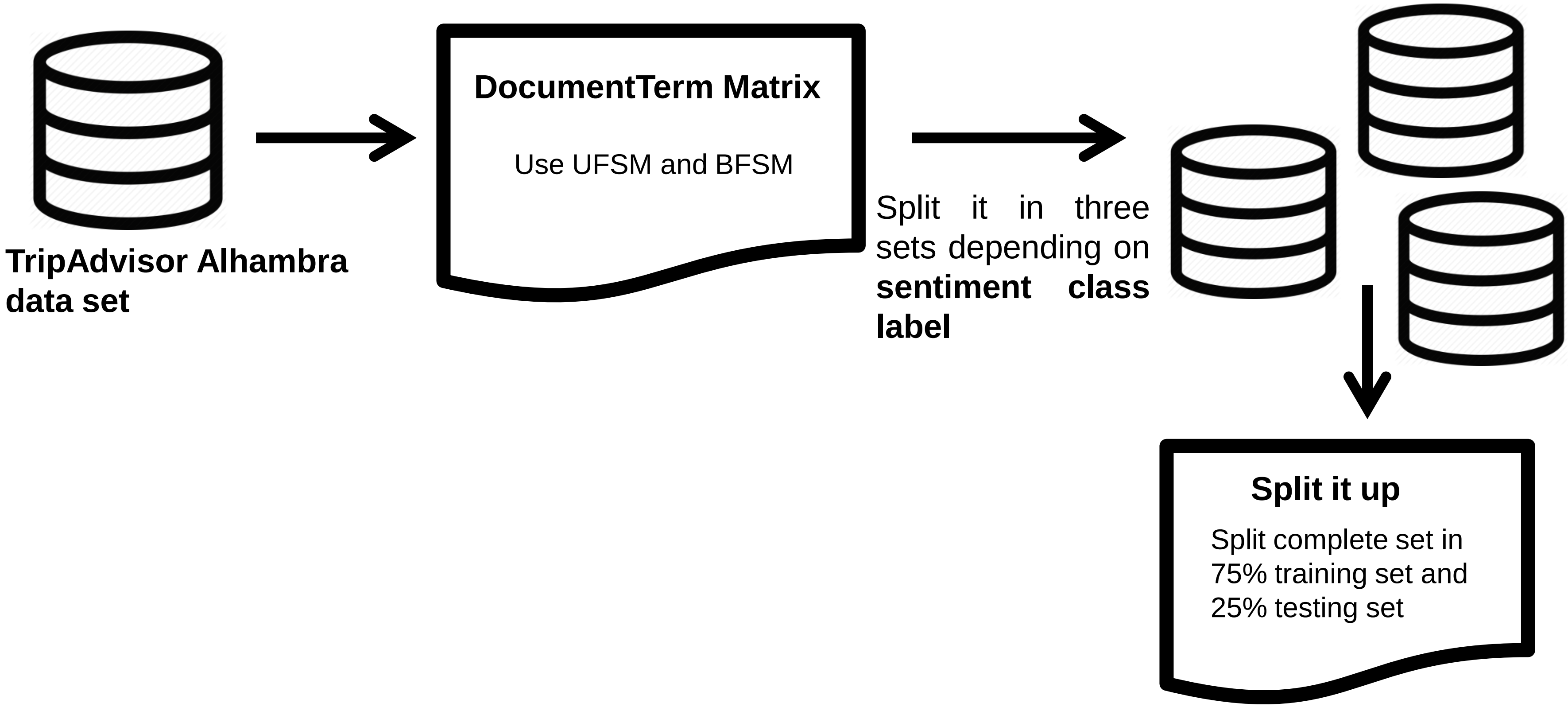
SentimentValue

SentimentValue ∩ SentimentCoreNLP

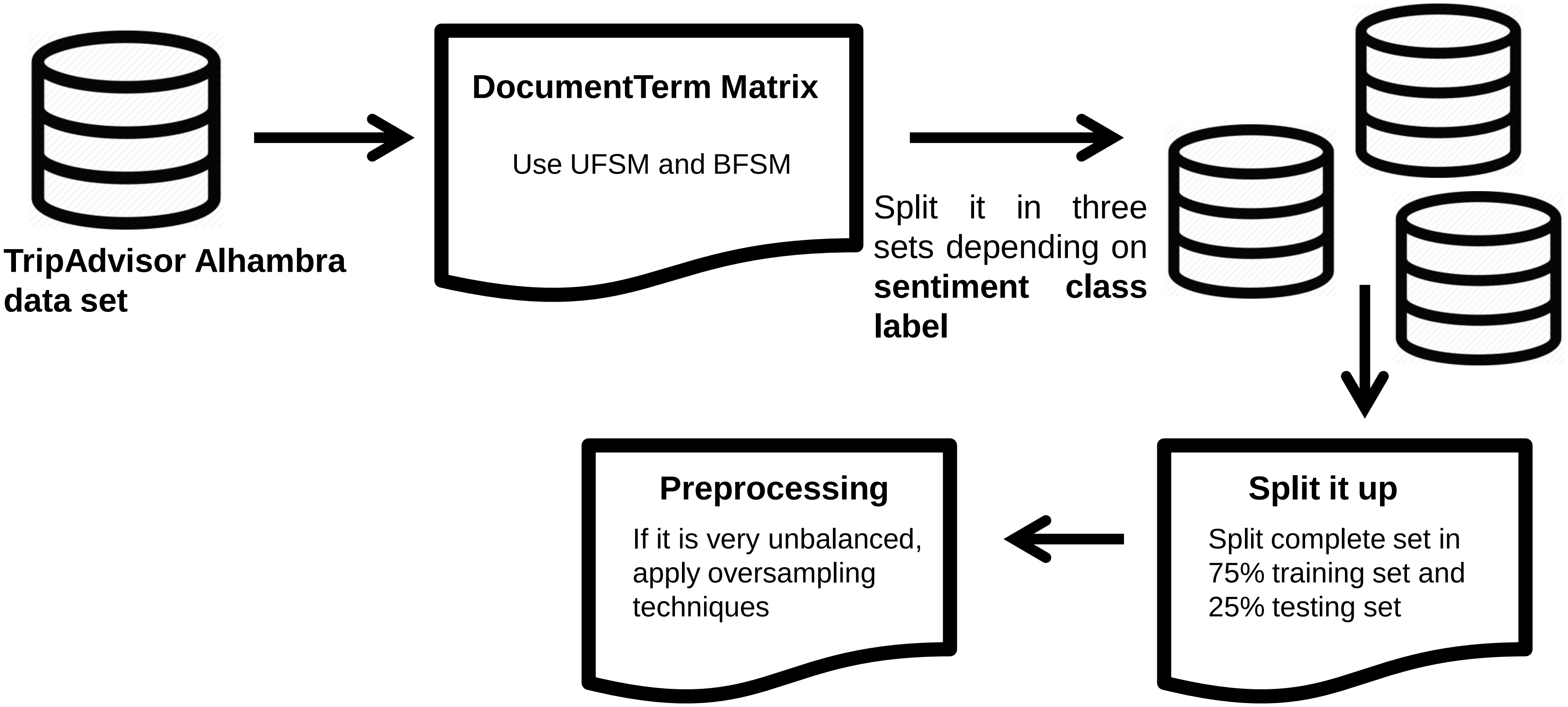
SentimentCoreNLP

	SentimentCoreNLP			
SentimentValue	positive	neutral	negative	Total
positive	4,049	1,071	2,508	7,628
neutral	51	32	260	343
negative	5	6	158	169
Total	4,105	1,109	2,926	8,140

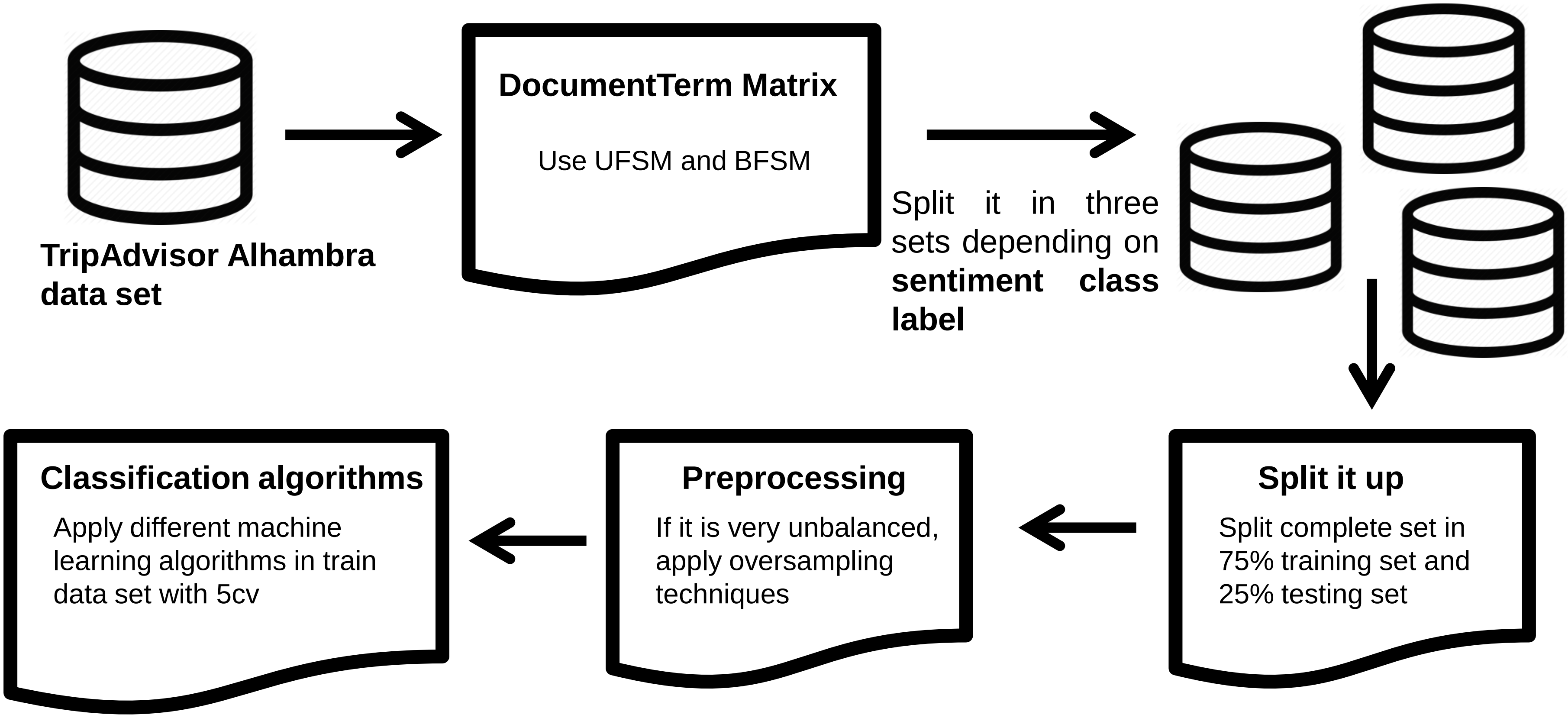
3. CLASSIFICATION ANALYSIS



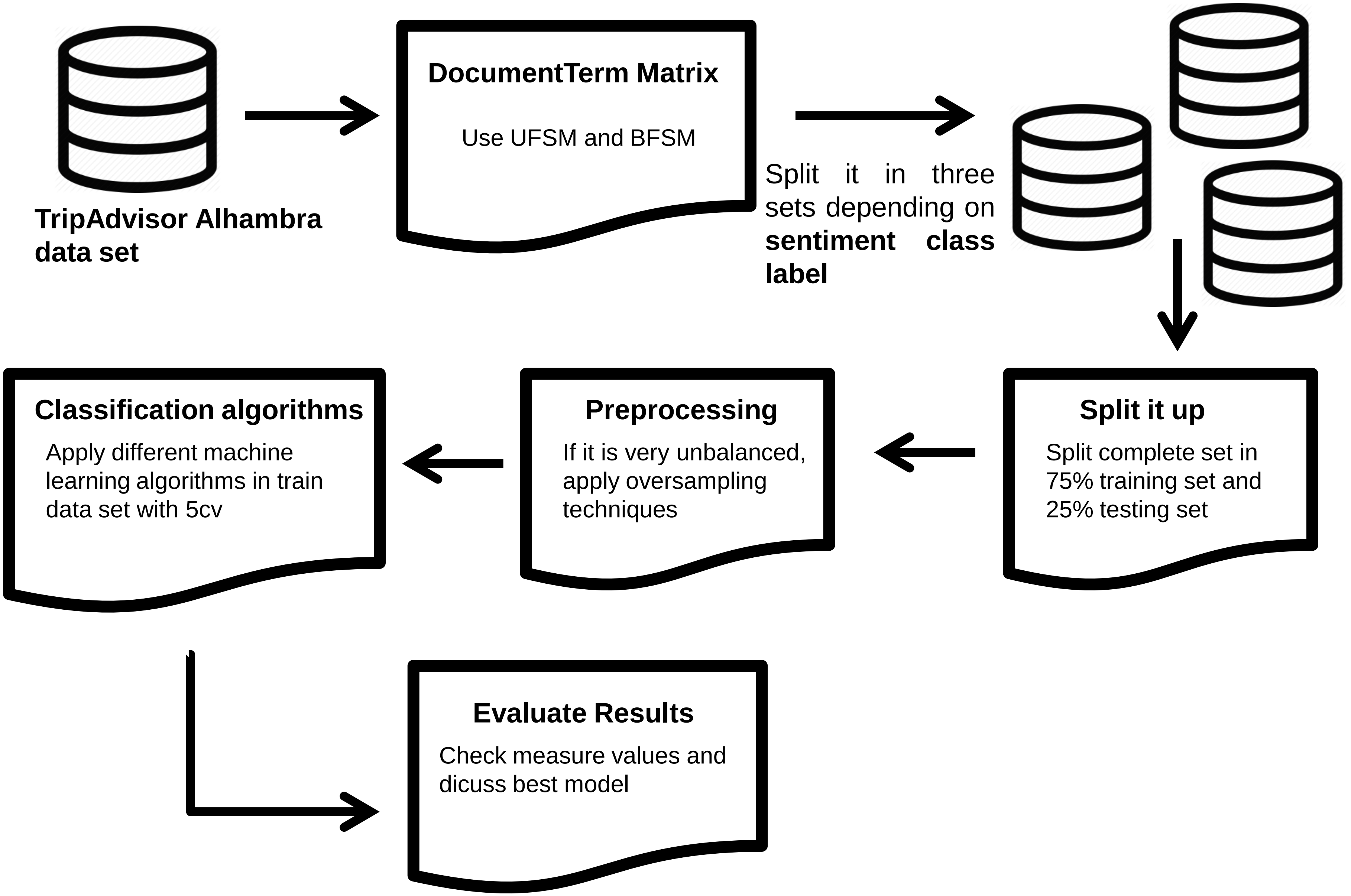
3. CLASSIFICATION ANALYSIS



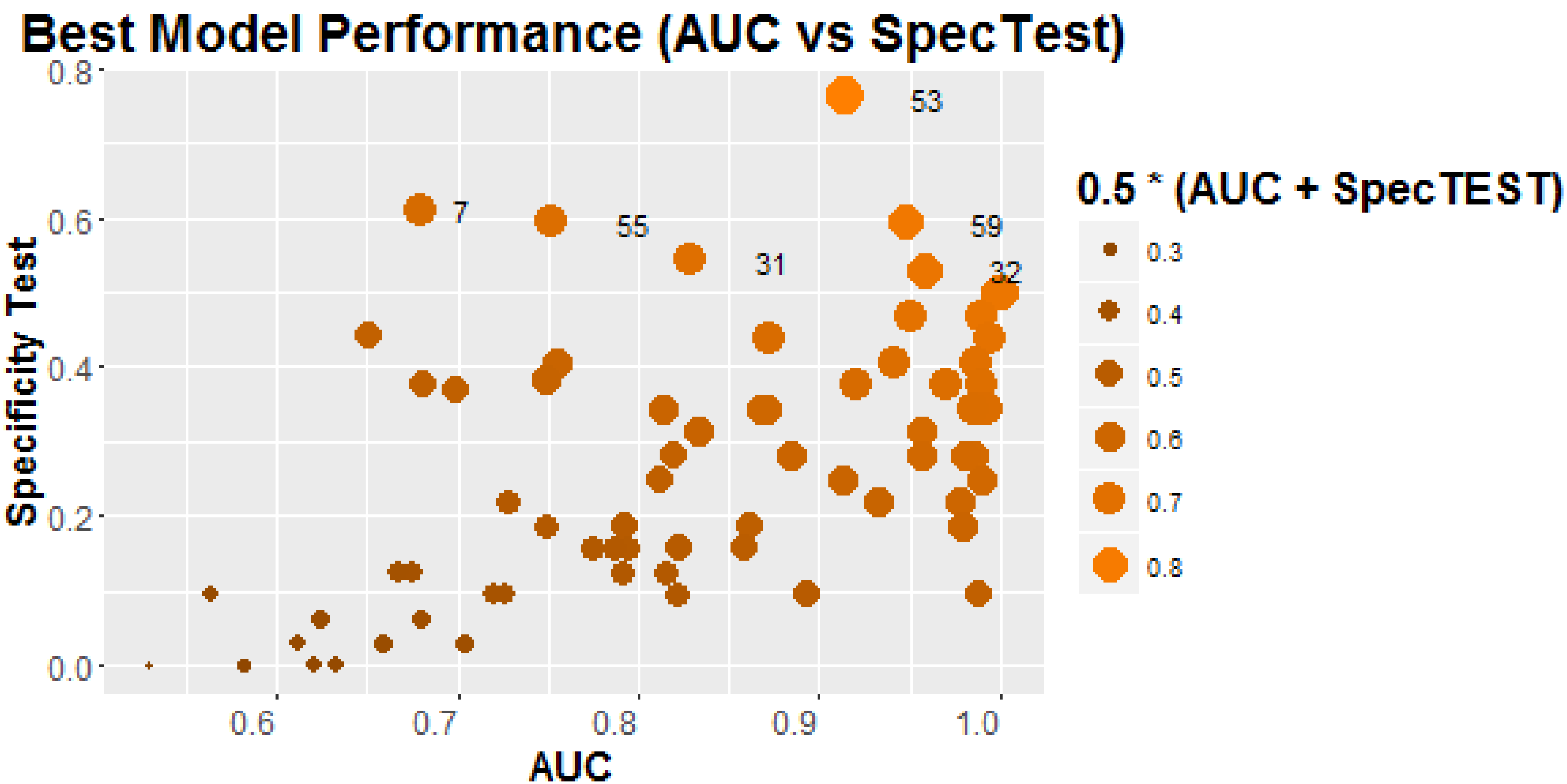
3. CLASSIFICATION ANALYSIS



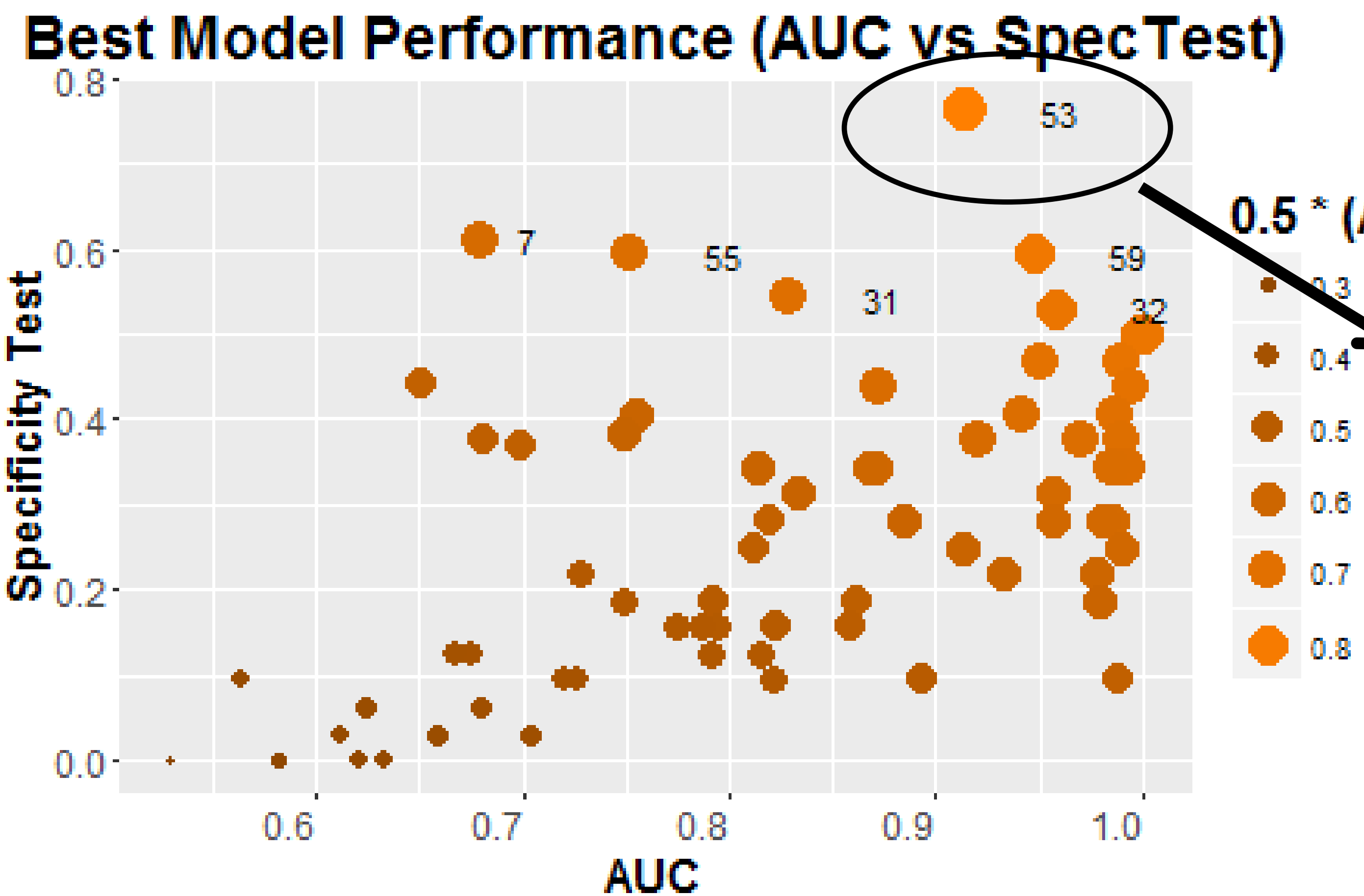
3. CLASSIFICATION ANALYSIS



3. CLASSIFICATION ANALYSIS



3. CLASSIFICATION ANALYSIS



XGBoost

SentimentValue

IR = 1

unigrams

4. SUBGROUP DISCOVERY

DISCOVERING INTERESTING PATTERNS IN NEGATIVE OPINIONS

4. SUBGROUP DISCOVERY

$$R : Cond \longrightarrow Target_{value}$$

negative

4. SUBGROUP DISCOVERY

INTERESTING PATTERNS

SentimentValue

$\cap$

SentimentCoreNLP

4. SUBGROUP DISCOVERY

INTERESTING PATTERNS

SentimentValue

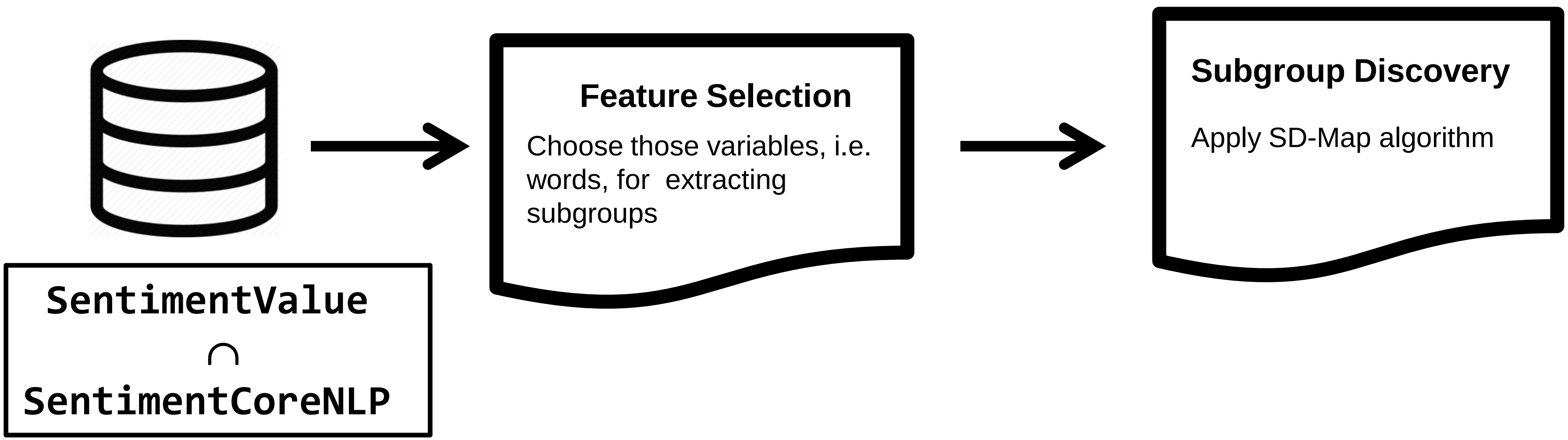
∩

SentimentCoreNLP

Feature Selection

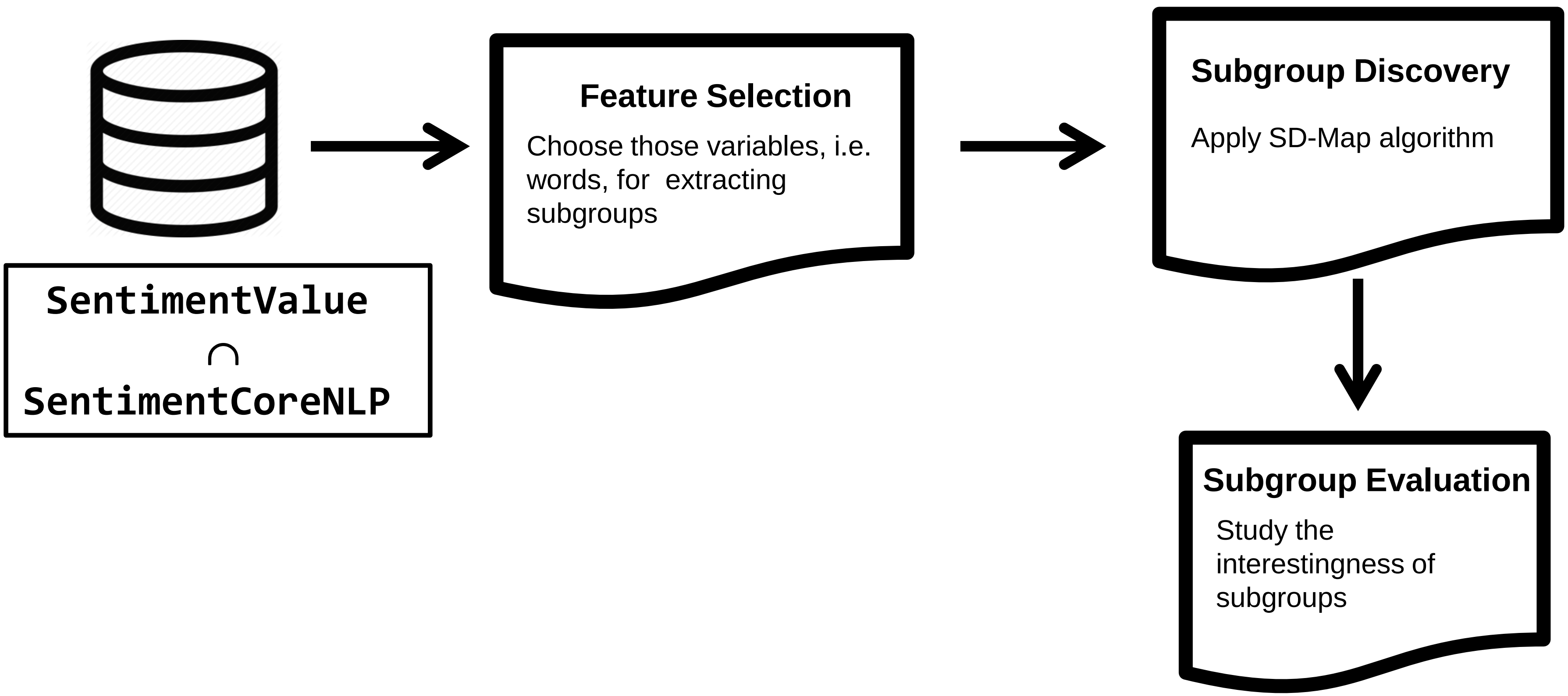
Choose those variables, i.e. words, for extracting subgroups

4. SUBGROUP DISCOVERY



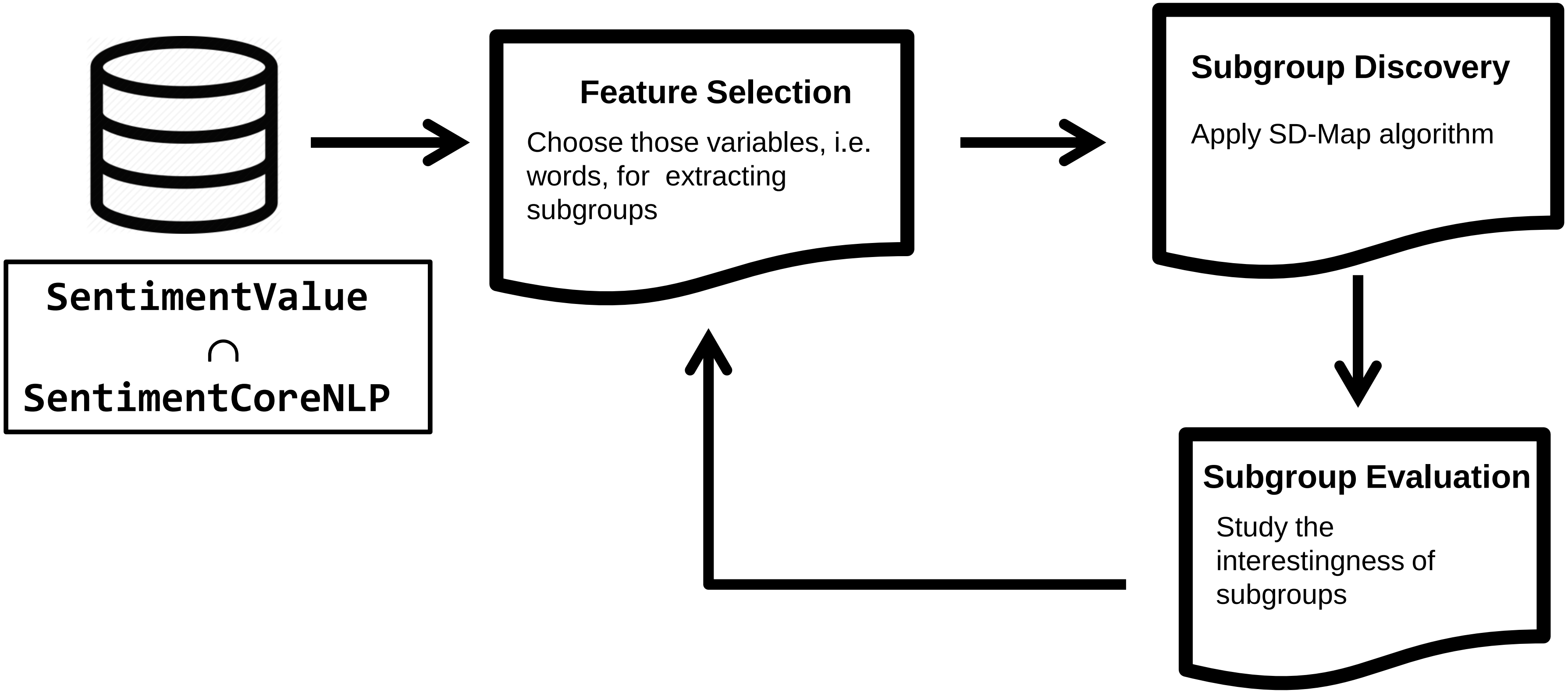
4. SUBGROUP DISCOVERY

INTERESTING PATTERNS



4. SUBGROUP DISCOVERY

INTERESTING PATTERNS



4. SUBGROUP DISCOVERY

INTERESTING PATTERNS

q_{BT}	p	n	**Description**
17.17	0.82	22	guard=1
16.21	0.81	21	terribl=1
12.29	0.68	19	rude=1
3.85	1	4	rude=1, guard=1
2.89	1	3	terribl=1, guard=1
2.77	0.5	6	babi=1
1.85	0.5	4	babi=1, strollX=1
1.6	0.06	64	strollX=1
30.33	0.46	71	staff=1
6.7	0.88	8	attitud=1
3.85	1	4	horribl=1
3.85	1	4	attitud=1, staff=1
3.85	1	4	horribl=1, staff=1
1.92	1	2	horribl=1, attitud=1, staff=1
30.33	0.14	284	queueX=1
27.06	0.06	1303	time=1
21.55	0.19	145	queueX=1, time=1
5.21	0.29	21	wheelchair=1
4.66	0.56	9	disabl=1
2.85	0.75	4	disabl=1, wheelchair=1
8.19	0.08	208	night=1
7.2	0.08	181	night=1, light=0
1.41	0.06	69	light=1
0.99	0.07	27	light=1, night=1
0.42	0.05	42	light=1, night=0
0.03	0.29	79	speakX=1
14.13	0.17	103	english=1
6.62	0.7	10	staff=1, english=1, speakX=1
7.62	0.8	10	email=1
5.47	0.43	14	confirm=1
3.85	1	4	email=1, confirm=1

Table 5.2: Subgroups generated from feature correlation.

CONCLUSIONS

5. CONCLUSION

In general…

1. Tourists are **satisfied** with the Alhambra

5. CONCLUSION

In general…

1. Tourists are **satisfied** with the Alhambra
2. The Alhambra should rethink the **ticket system**

5. CONCLUSION

More precisely…

1. Suitable TripAdvisor Alhambra **data set** for applying sentiment analysis

5. CONCLUSION

More precisely…

1. Suitable TripAdvisor Alhambra **data set** for applying sentiment analysis

2. **Low correlation** between human and machine sentiment

5. CONCLUSION

More precisely…

1. Suitable TripAdvisor Alhambra **data set** for applying sentiment analysis

2. **Low correlation** between human and machine sentiment

3. **Good** model performance in the **classification** analysis

5. CONCLUSION

More precisely…

1. Suitable TripAdvisor Alhambra **data set** for applying sentiment analysis

2. **Low correlation** between human and machine sentiment

3. **Good** model performance in the **classification** analysis

4. **Poor** analysis of **subgroup discovery**

5. CONCLUSION

FURTHER WORK

5. CONCLUSION

Next months…

1. Talk to **Patronato** about this project

5. CONCLUSION

Next months…

1. Talk to **Patronato** about this project

2. Work on the **sentiment labels**

5. CONCLUSION

Next months…

1. Talk to **Patronato** about this project
2. Work on the **sentiment labels**
3. Apply **aspect-sentiment** algorithms

5. CONCLUSION

Next months…

1. Talk to **Patronato** about this project
2. Work on the **sentiment labels**
3. Apply **aspect-sentiment** algorithms
4. Improve **subgroup discovery** method

5. CONCLUSION

Next months…

1. Talk to **Patronato** about this project
2. Work on the **sentiment labels**
3. Apply **aspect-sentiment** algorithms
4. Improve **subgroup discovery** method
5. Consider **neutral** sentiment

5. CONCLUSION

Next months…

1. Talk to **Patronato** about this project
2. Work on the **sentiment labels**
3. Apply **aspect-sentiment** algorithms
4. Improve **subgroup discovery** method
5. Consider **neutral** sentiment
6. Develop **tool** for touristic attraction managers

THANKS!

MASTER'S DEGREE THESIS

Sentiment Analysis For Touristic Attractions: A Case Study On The Alhambra

Ana Valdivia García

19 September 2016

Department of Computer Science and Artificial Intelligence

Universidad de Granada