

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



ugr

Universidad
de Granada

CONTENTS

- 1. INTRODUCTION**
- 2. DATA**
- 3. CLASSIFICATION ANALYSIS**
- 4. SUBGROUP DISCOVERY**
- 5. CONCLUSION**

1. INTRODUCTION

BACKGROUND AND MOTIVATION

1. INTRODUCTION

BACKGROUND AND MOTIVATION

Tourism



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

1. INTRODUCTION

BACKGROUND AND MOTIVATION

Web 2.0



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

1. INTRODUCTION

BACKGROUND AND MOTIVATION

Data



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

1. INTRODUCTION

BACKGROUND AND MOTIVATION

What tourists think about touristic attractions?

1. INTRODUCTION

OBJECTIVES

First approach to sentiment analysis
into touristic attractions domain

1. INTRODUCTION

OBJECTIVES

More precisely...

1. Download and **analyse** reviews

1. INTRODUCTION

OBJECTIVES

More precisely...

1. **Download** and **analyse** reviews
2. **Study** correlation between **human** and **machine sentiment**

1. INTRODUCTION

OBJECTIVES

More precisely...

1. **Download** and **analyse** reviews
2. **Study** correlation between **human** and **machine sentiment**
3. **Classify** opinions

1. INTRODUCTION

OBJECTIVES

More precisely...

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

2. DATA

Where do we get the data?

2. DATA

TRIP ADVISOR



TripAdvisor logo

2. DATA

THE ALHAMBRA



The Alhambra

2. DATA

SCRAPPING

TFM Alhambra Experiment x W Feature hashing - Wikiped x The Alhambra, Granada x Ana

← → ↺ 🏠 TripAdvisor LLC [US] https://www.tripadvisor.co.uk/Attraction_Review-g187441-d191078-Reviews-The_Alhambra-Granada_Province_of_Granada_Andalucia.html 🔍 ☆ 40 ☰

Aplicaciones Gmail Facebook correo UGR UGR SWAD Comedores UGR Colors in R Library Genesis TFM Diccionarios de Español TableIDbf8163208fcf » Otros marcadores

The Alhambra

🟢🟢🟢🟢🟢 20,832 Reviews #1 of 194 things to do in Granada


Neighbourhood: Realejo - San Matias Architectural Buildings, Historic Sites, Sights & Landmarks, More

As featured in 3 Days in Granada and 1 other guide

2016 TRAVELLERS' CHOICE tripadvisor

Overview Tours & Tickets Reviews (20,832) Q&A (191) Location

Book a Tour Save Write a Review




All visitor photos (13625)

Book In Advance

Granada Highlights: Guided Walking Tour	€55* and up	More Info
Guided Night Alhambra Tour with Nazrid Palaces	€75* and up	More Info
Skip the Line: Alhambra and Generalife Gardens Half-Day Tour	€55* and up	More Info

See More Tours & Experiences



Address: Palacio de Carlos V, Granada, Spain
Phone Number: 958 027 971
[Website](#)

Improve this listing

Recommended length of visit: More than 3 hours

2. DATA

SCRAPPING


The Alhambra, Granada - x

← → ↺ 🏠 TripAdvisor LLC [US] https://www.tripadvisor.co.uk/Attraction_Review-g187441-d191078-Reviews-The_Alhambra-Granada_Province_of_Granada_Andalucia.html 🔍 ☆ 30 ≡

Aplicaciones Gmail Facebook correo UGR UGR SWAD Comedores UGR Colors in R Library Genesis TFM Diccionarios de Españ TableIDbf8163208fcf >> Otros marcadores

Overview Tours & Tickets **Reviews (20,803)** Q&A (191) Location

Book a Tour Save Write a Review




surf_golf
Guaynabo, Puerto Rico
Level 6 Contributor
158 reviews
33 attraction reviews
39 helpful votes

"Amazing Experience!" NEW

★★★★★ Reviewed yesterday via mobile

A must visit in Granada! An amazing journey through history makes you truly appreciate this gem. We really enjoyed our visit. Plenty of walking required.

Helpful? 1 Thank surf_golf Report




Joanne F
Washington, DC
Level 5 Contributor
69 reviews
25 attraction reviews
22 helpful votes

"Quite Spectacular" NEW

★★★★★ Reviewed yesterday

The Nasrid Palaces of the Alhambra are just amazing works of architecture and design. It is a wonderful place to visit. Our regret was not having a tour guide or getting an audio tour for this complex. I had so many questions about the designs that went unanswered.



Helpful? 1 Thank Joanne F Report

21 friends have been to Granada & nearby towns

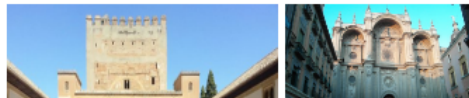
Which Granada hotels are on sale?

dd/mm/yyyy dd/mm/yyyy

See hotels

Sponsored links *

Don't miss the best of Granada



2. DATA

STRUCTURE

8140 instances



**TripAdvisorAlhambra
data set**

2. DATA

STRUCTURE



**TripAdvisorAlhambra
data set**

8140 instances

10 features

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

2. DATA

STRUCTURE



**TripAdvisorAlhambra
data set**

8140 instances

10 features

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

2 class labels

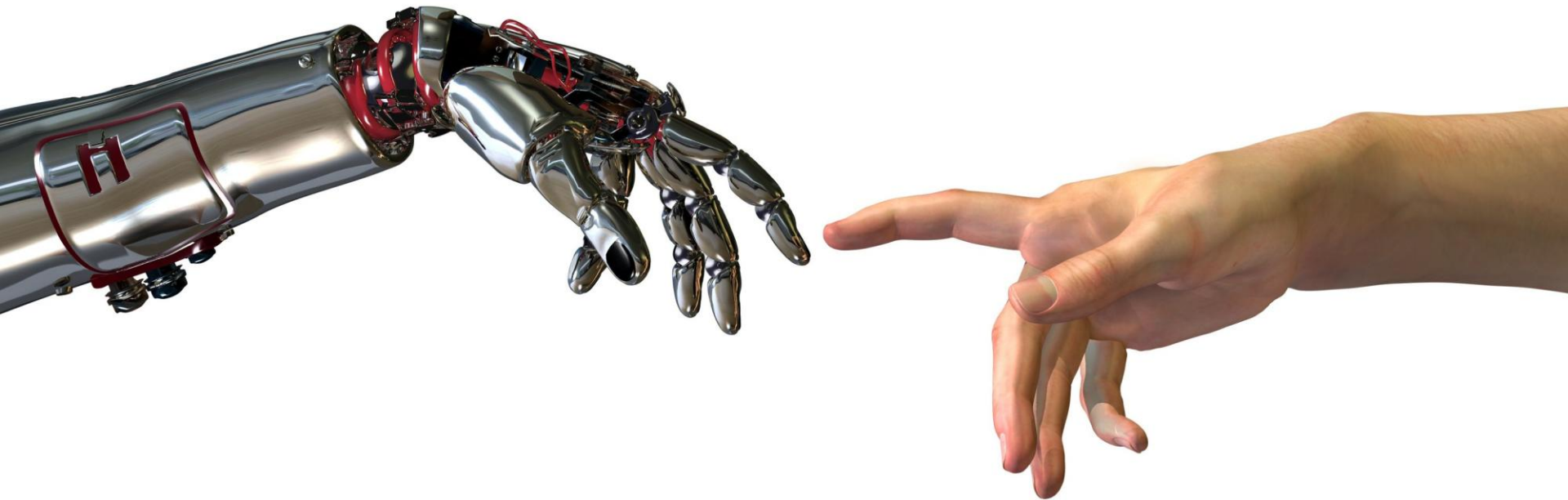
2. DATA

STRUCTURE

Class labels

SentimentCoreNLP

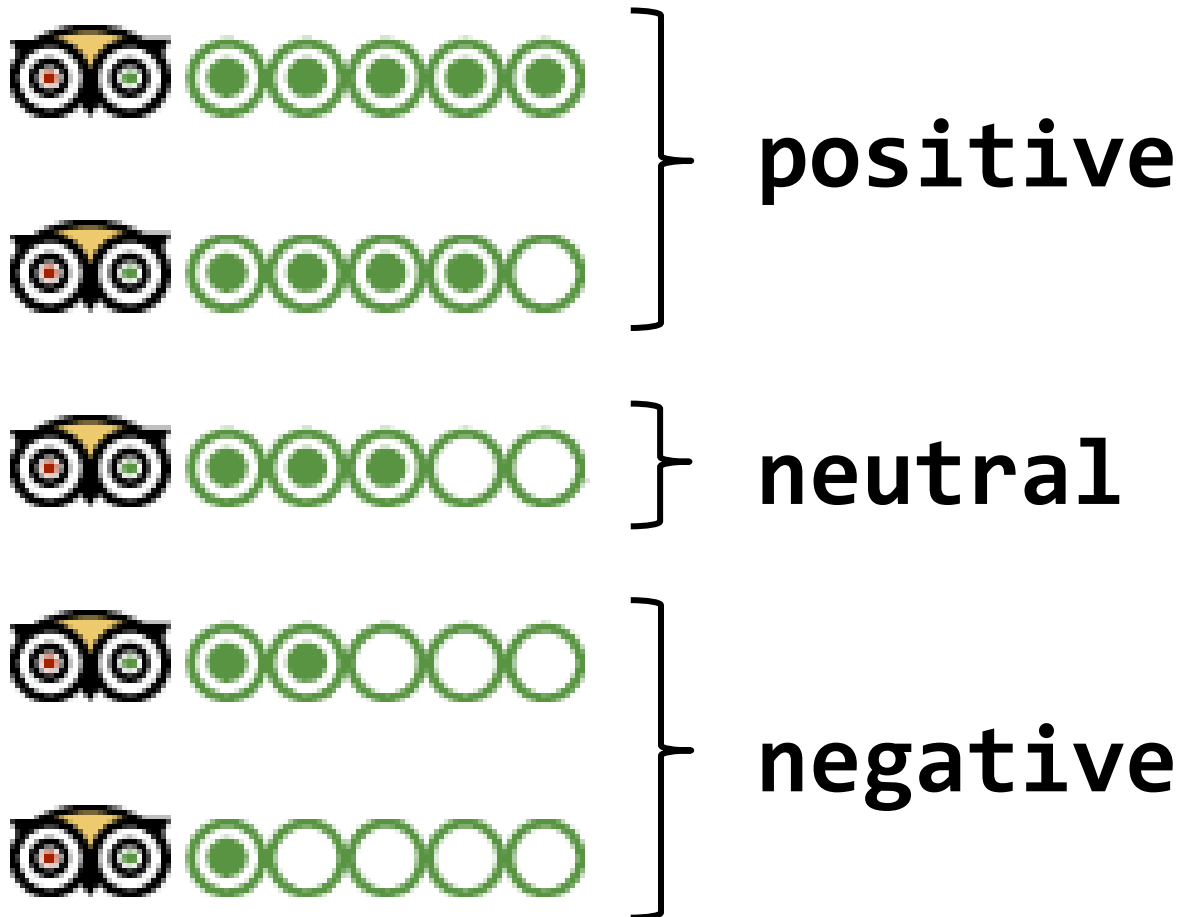
SentimentValue



2. DATA

STRUCTURE

SentimentValue



2. DATA

STRUCTURE

reviews

SentimentCoreNLP



Lindorian
Bottom
Level 4 Contributor
24 reviews
10 attraction reviews
13 helpful votes

"Very Moorish ;-)"

Reviewed 30 June 2016

Judging by the queues pre booking a must (which we did)
Get in the 'lined 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 plaster 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 a maximum of €18 so no rush to get out early.

Visited June 2016

Less +

Helpful? 1 Thank Lindorian

Report

Ask Lindorian about The Alhambra

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

2. DATA

STRUCTURE

reviews

SentimentCoreNLP



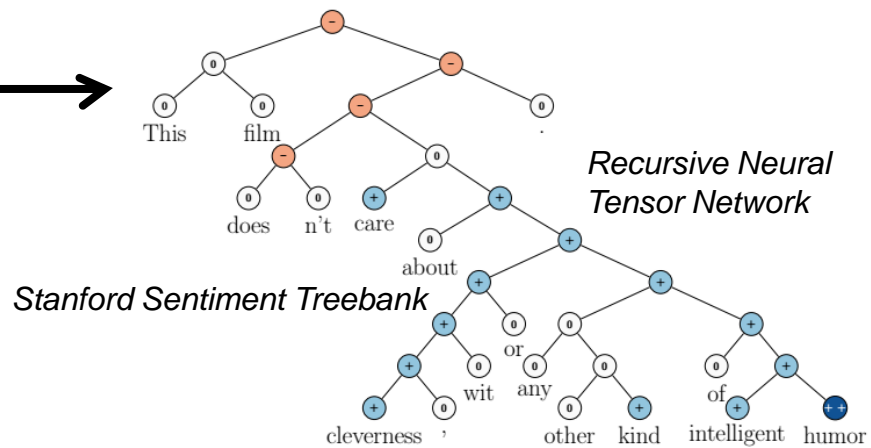
CoreNLP toolkit

reviews

SentimentCoreNLP



CoreNLP toolkit



2. DATA

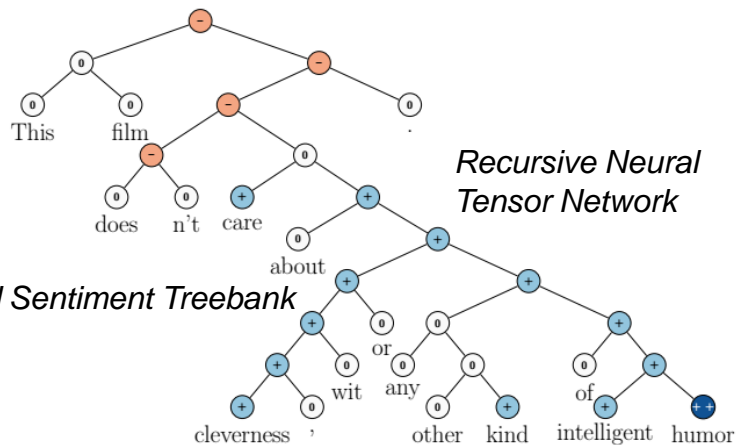
STRUCTURE

reviews

SentimentCoreNLP



CoreNLP toolkit



Σ

2. DATA

STRUCTURE

reviews

SentimentCoreNLP



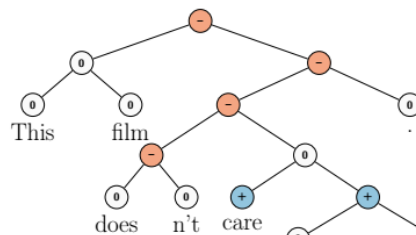
positive

neutral

negative

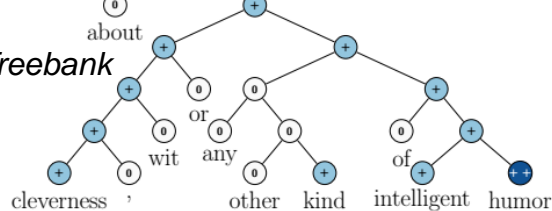
Σ

CoreNLP toolkit



Recursive Neural
Tensor Network

Stanford Sentiment Treebank



3. CLASSIFICATION ANALYSIS

ANALYSIS OF CORRELATION BETWEEN CLASS LABELS

3. CLASSIFICATION ANALYSIS

CORRELATION CLASS LABELS

SentimentValue	SentimentCoreNLP			Total
	positive	neutral	negative	
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

SentimentValue	SentimentCoreNLP			Total
	positive	neutral	negative	
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

3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS

3. CLASSIFICATION ANALYSIS

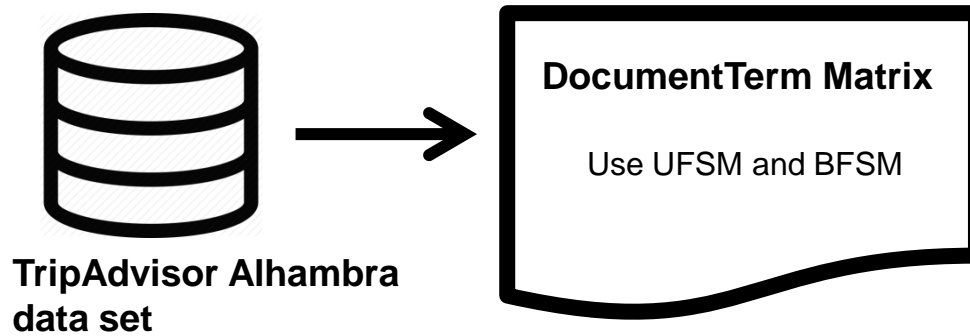
CLASSIFICATION ANALYSIS



**TripAdvisor Alhambra
data set**

3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS



CLASSIFICATION ANALYSIS



TripAdvisor Alhambra data set

DocumentTerm Matrix

Use UFSM and BFSM

UFSM

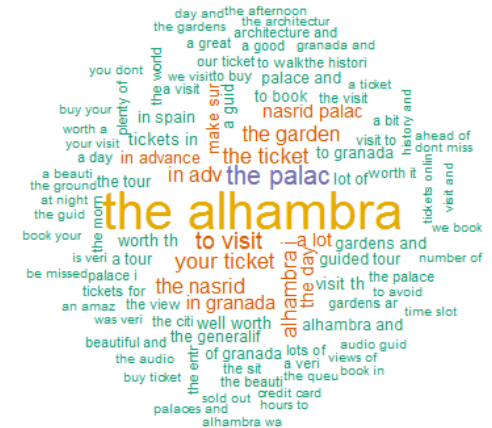


positive



negative

positive



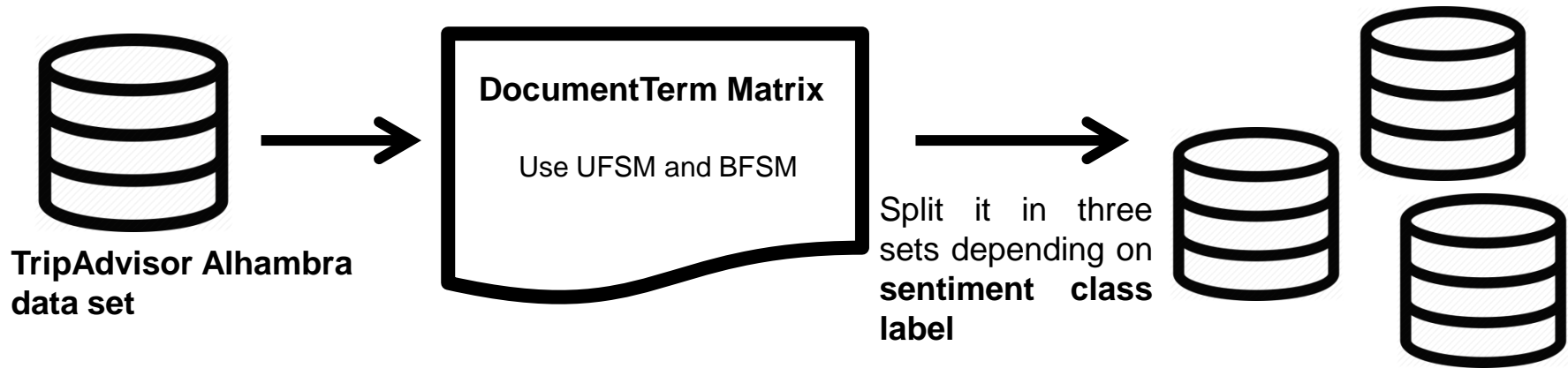
BFSM



negative

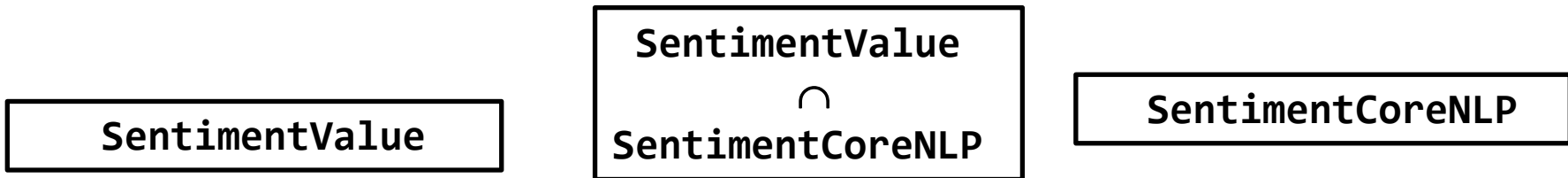
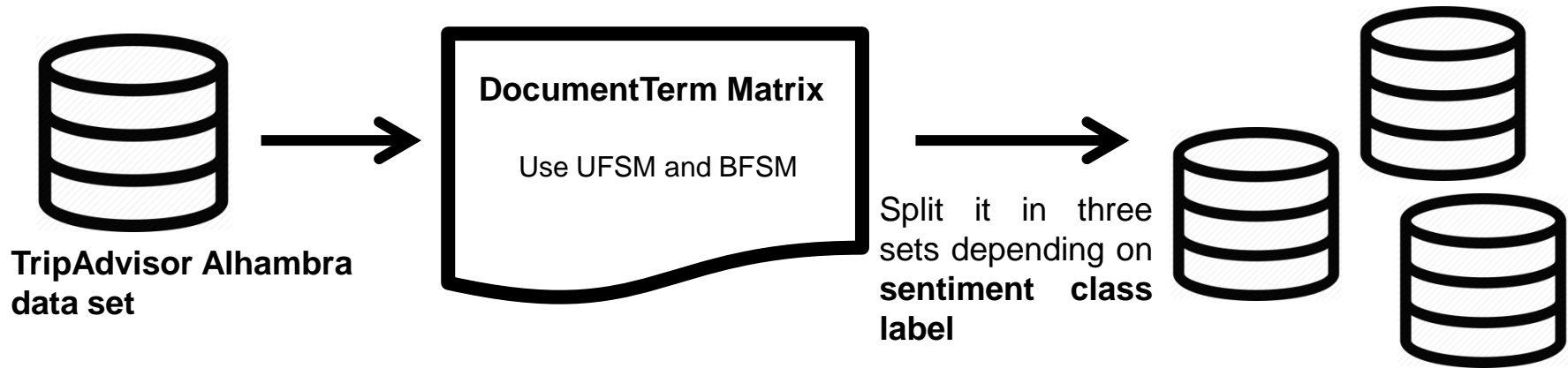
3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS



3. CLASSIFICATION ANALYSIS

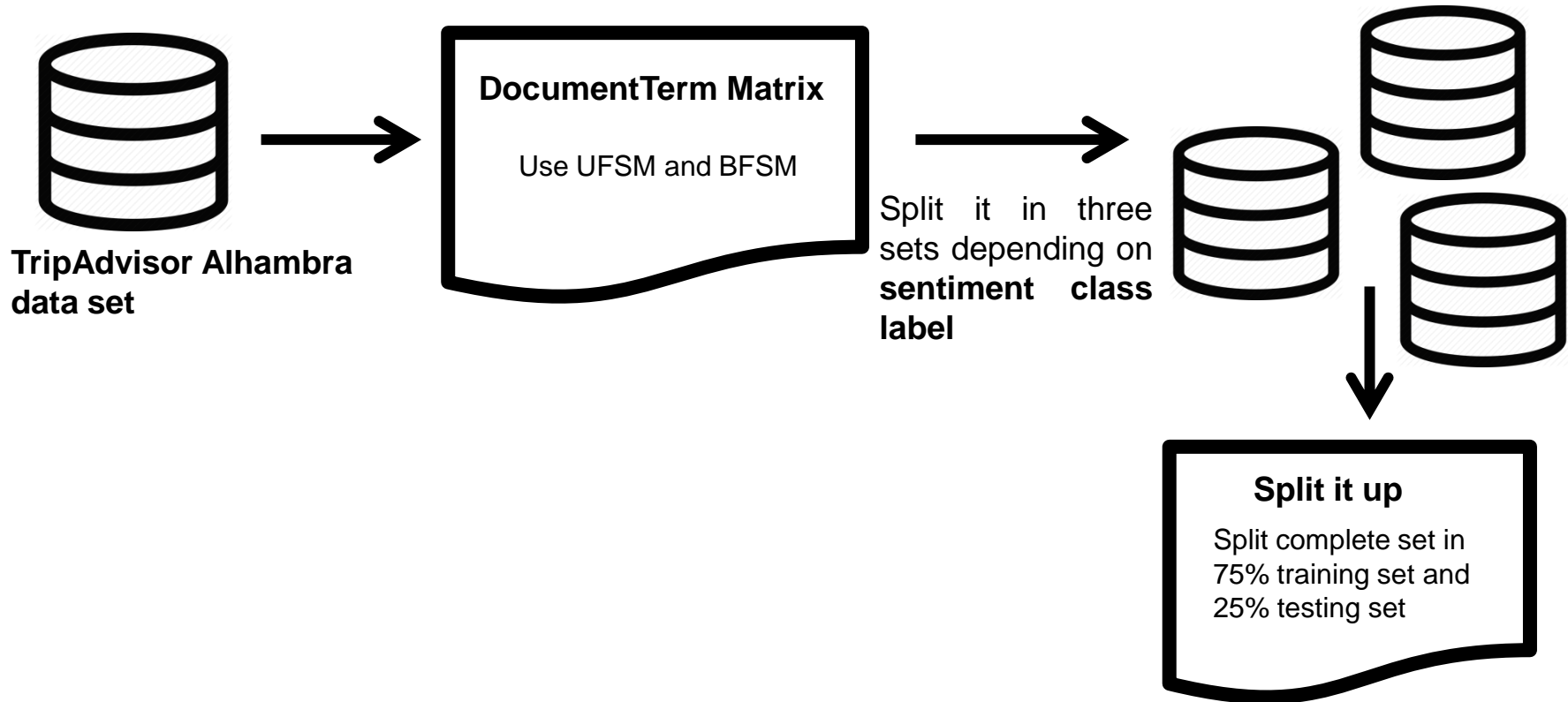
CLASSIFICATION ANALYSIS



SentimentValue	SentimentCoreNLP			Total
	positive	neutral	negative	
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

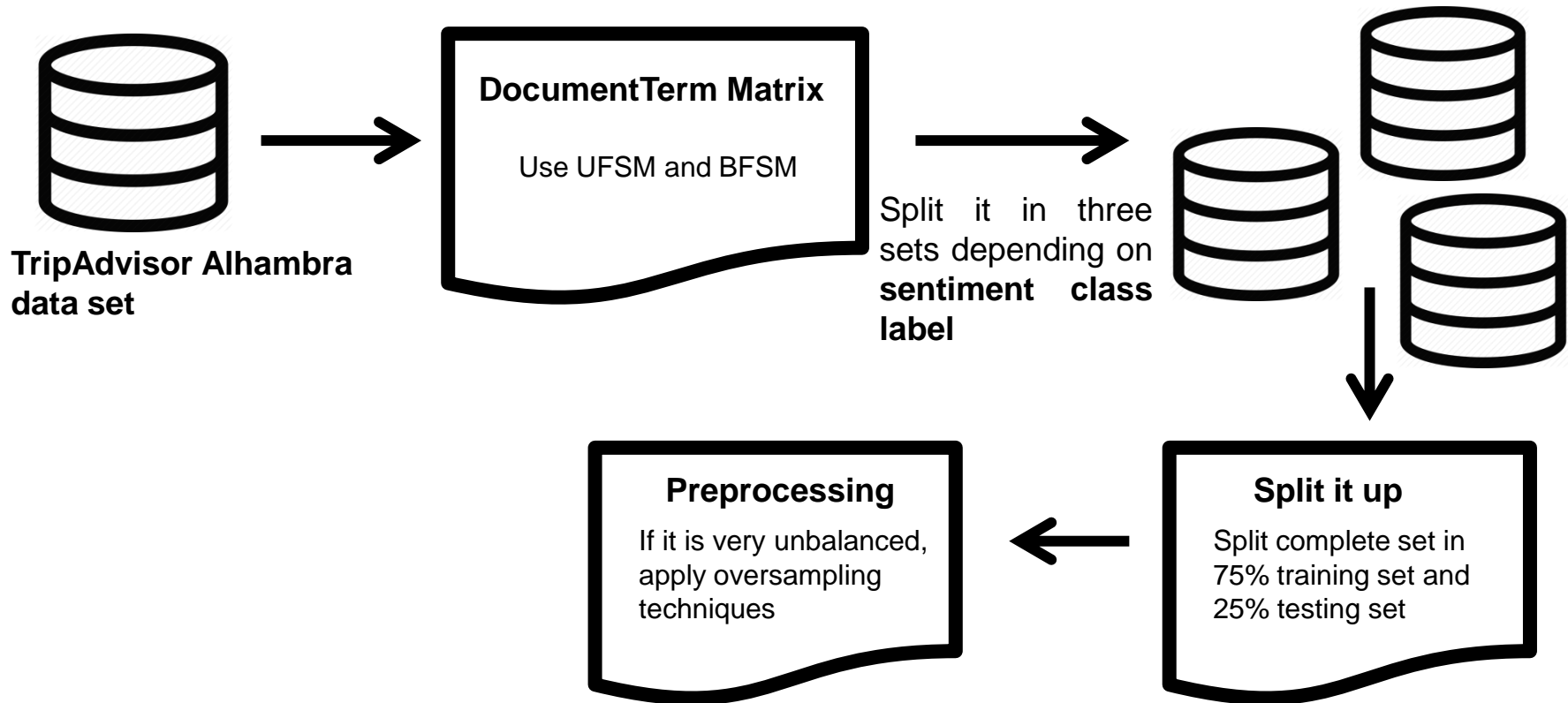
3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS



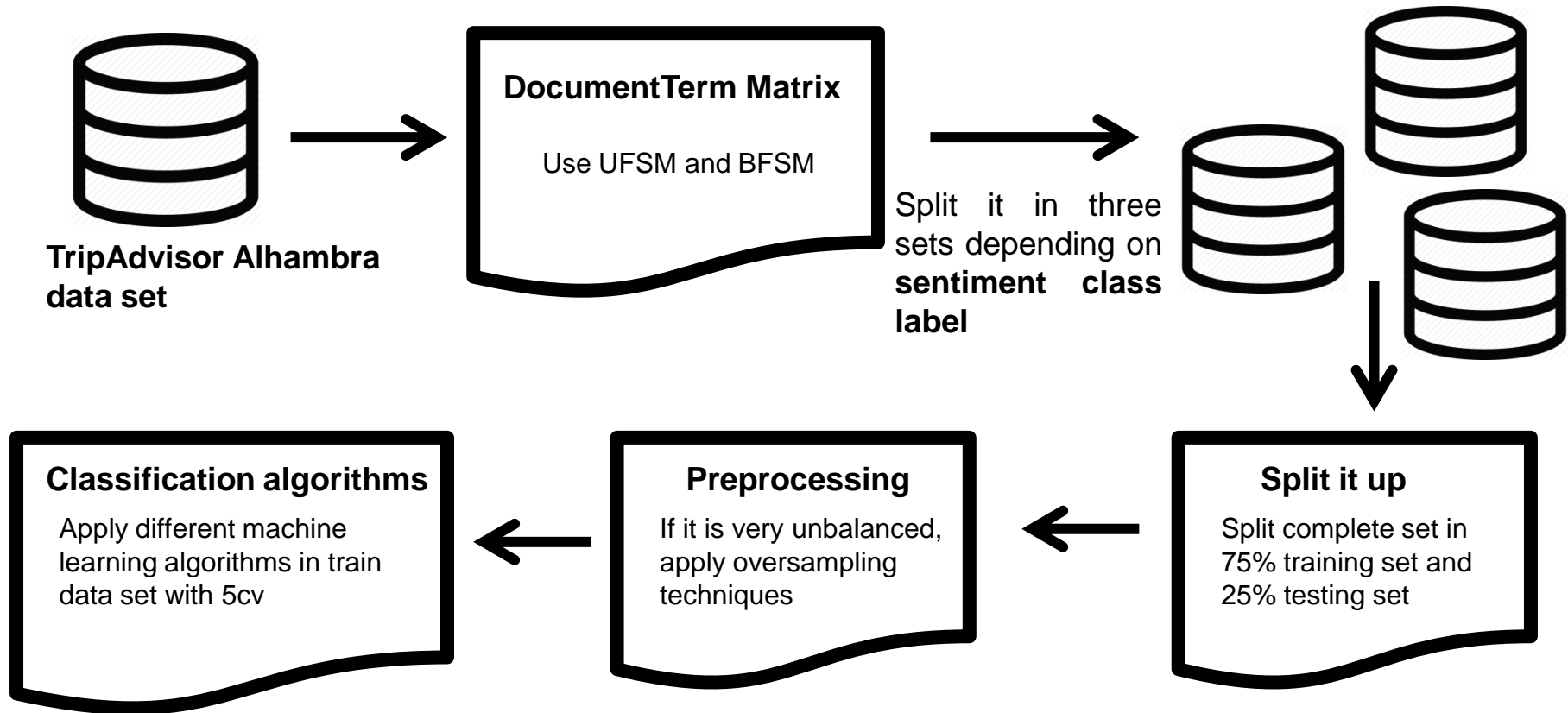
3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS



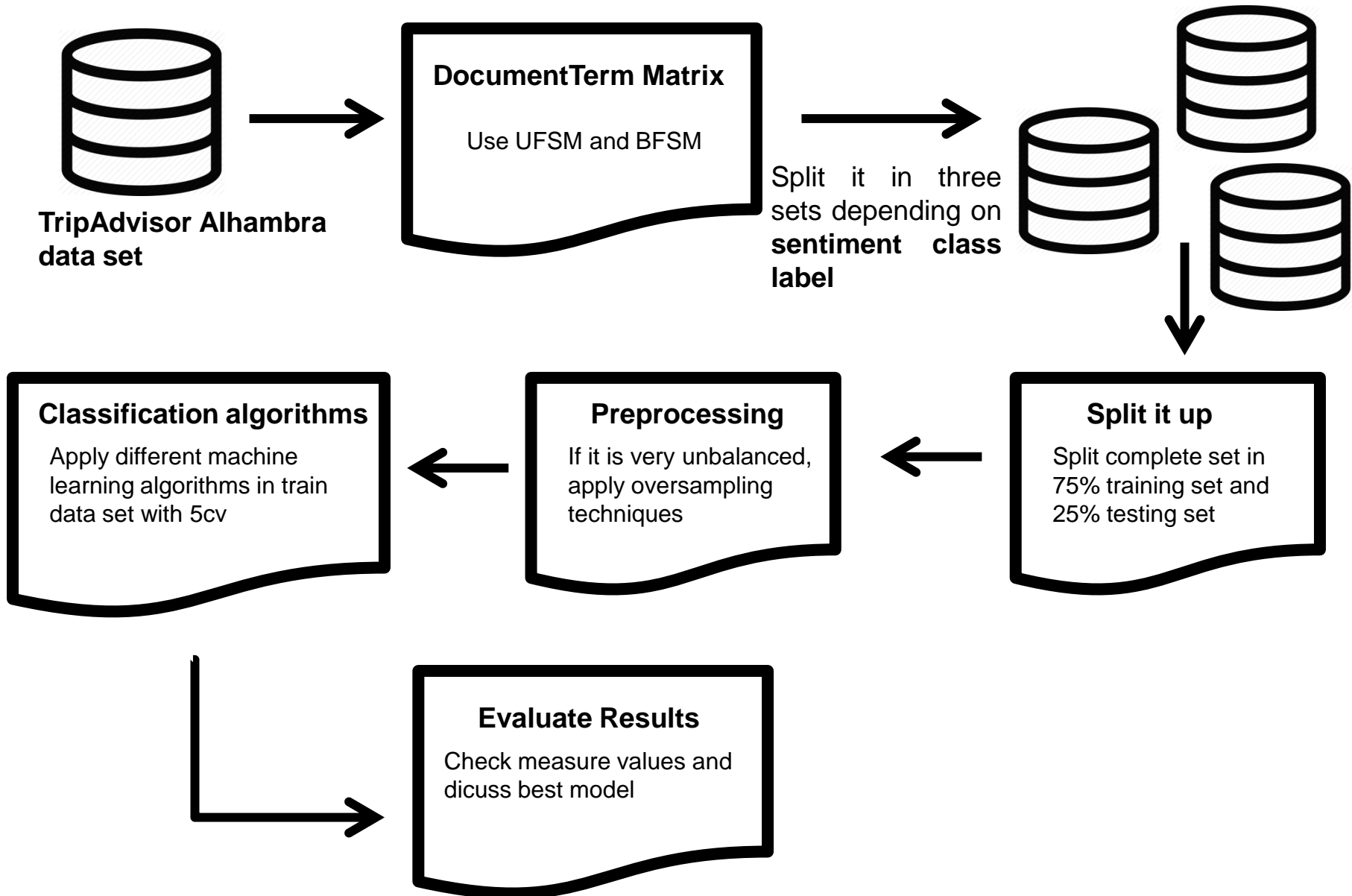
3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS



3. CLASSIFICATION ANALYSIS

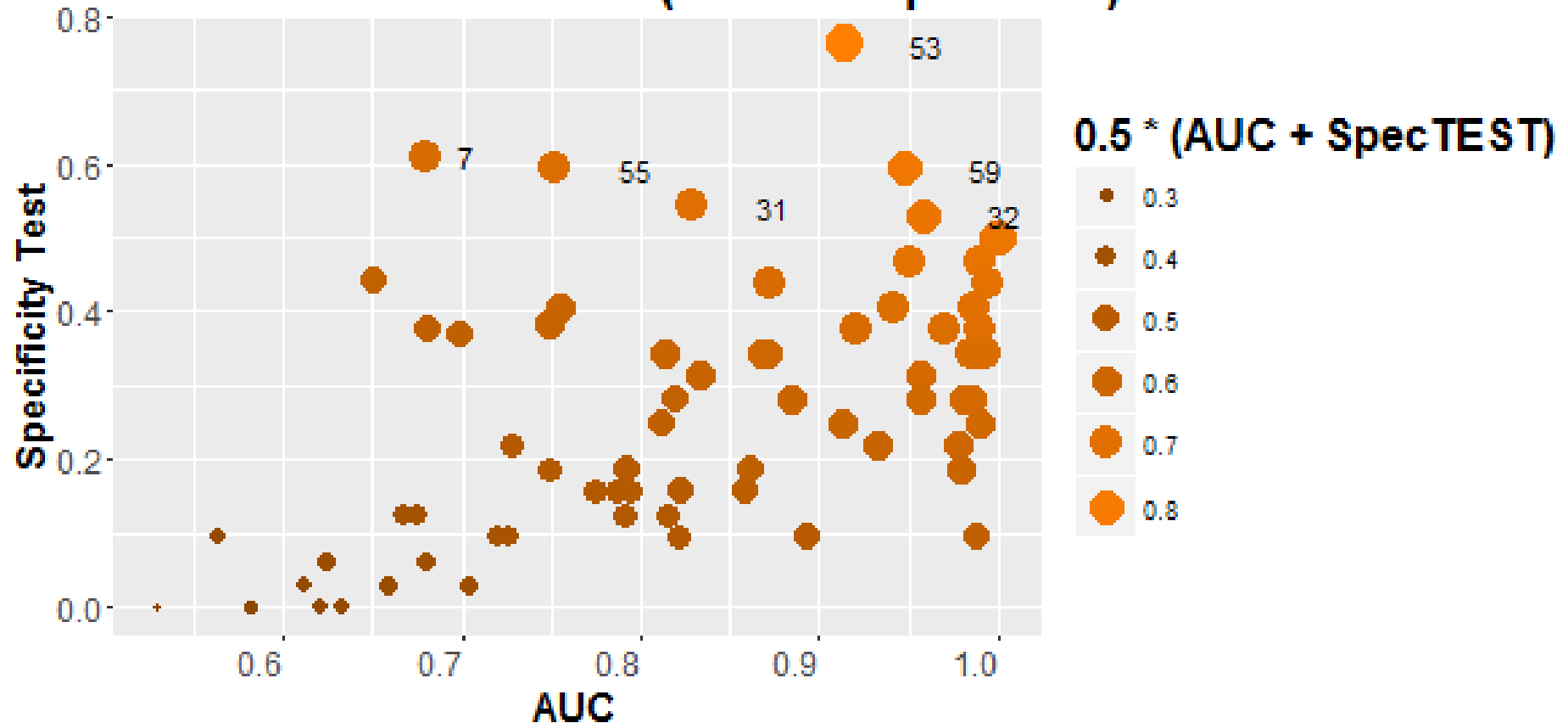
CLASSIFICATION ANALYSIS



3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS

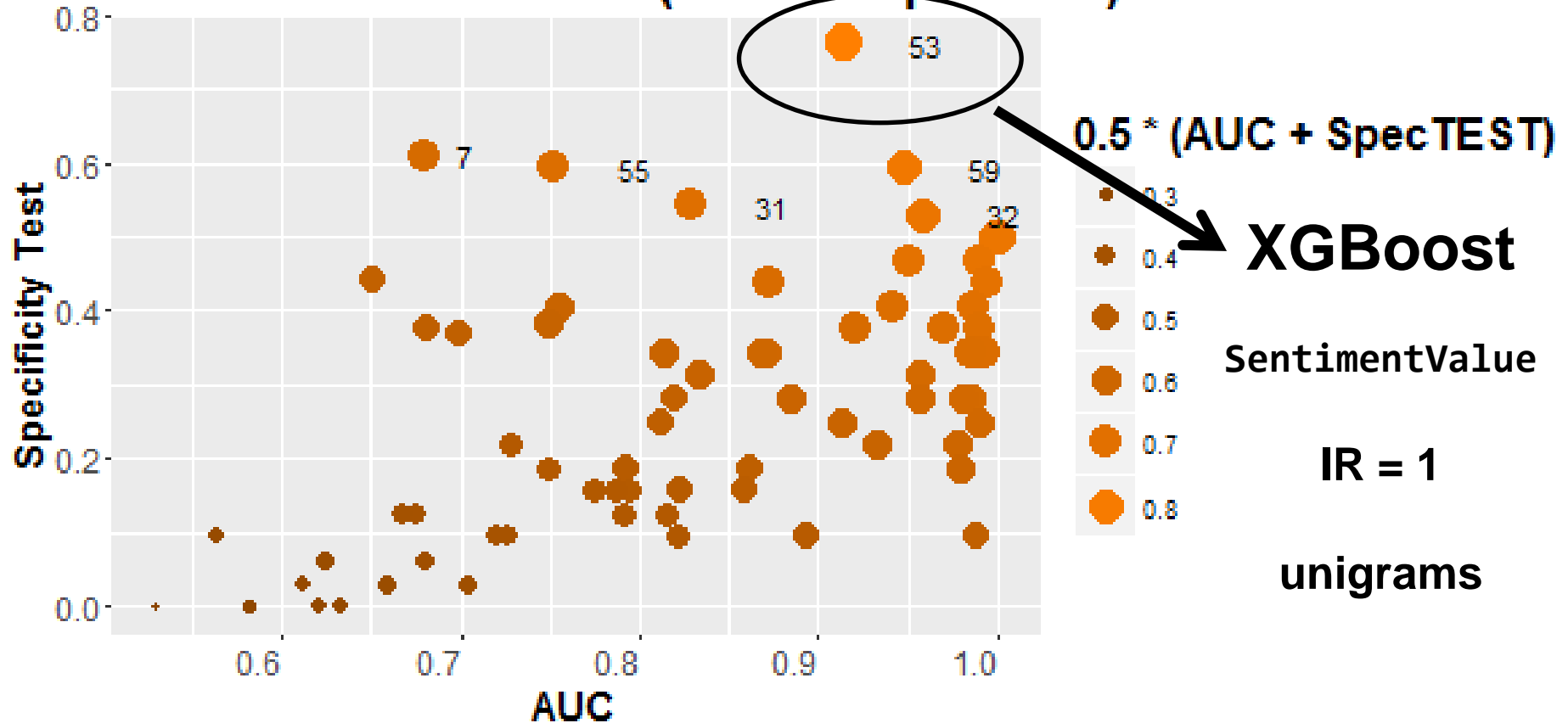
Best Model Performance (AUC vs SpecTest)



3. CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS

Best Model Performance (AUC vs SpecTest)



4. SUBGROUP DISCOVERY

**DISCOVERING INTERESTING PATTERNS IN
NEGATIVE OPINIONS**

4. SUBGROUP DISCOVERY

INTERESTING PATTERNS

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

negative

4. SUBGROUP DISCOVERY

INTERESTING PATTERNS



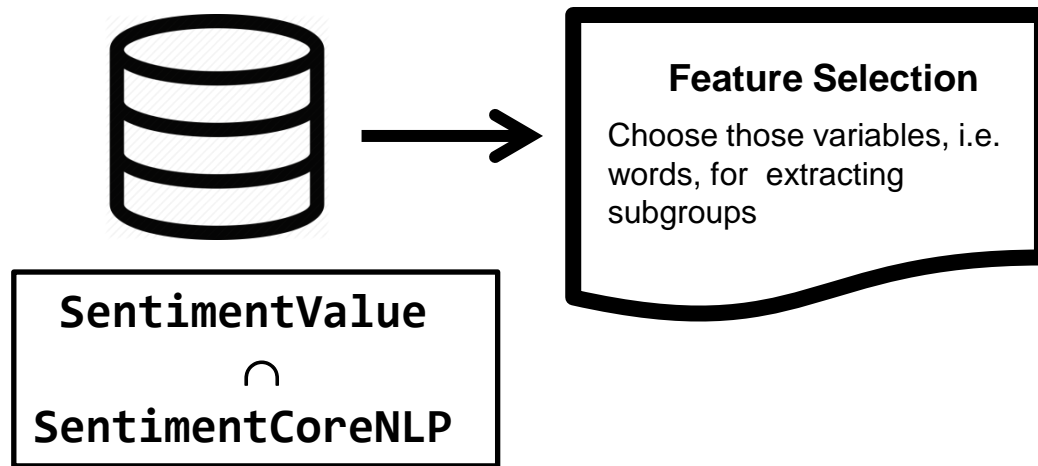
SentimentValue



SentimentCoreNLP

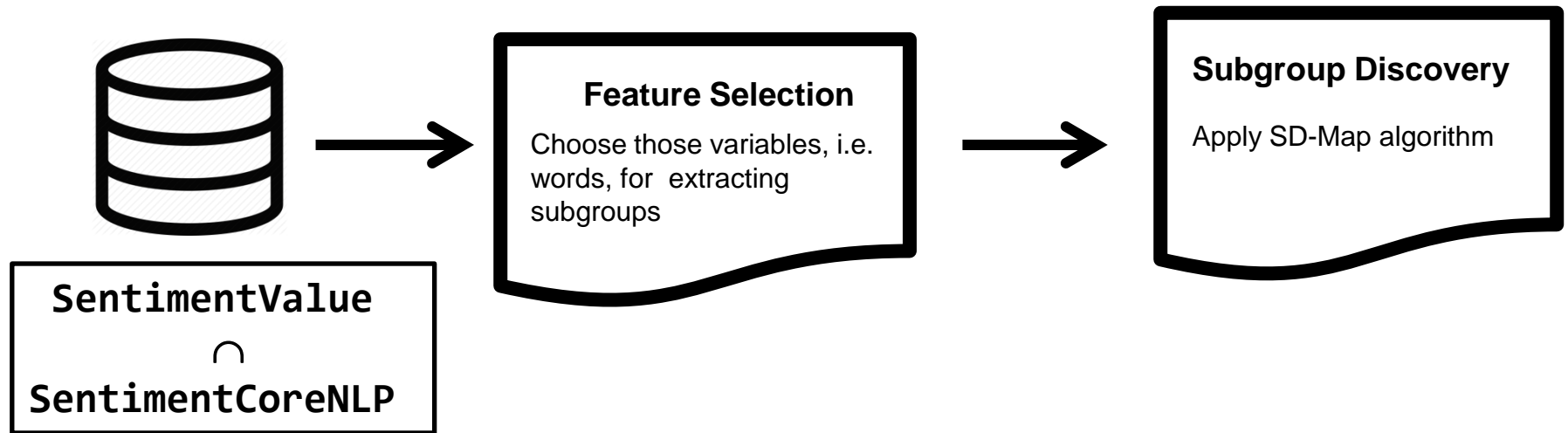
4. SUBGROUP DISCOVERY

INTERESTING PATTERNS



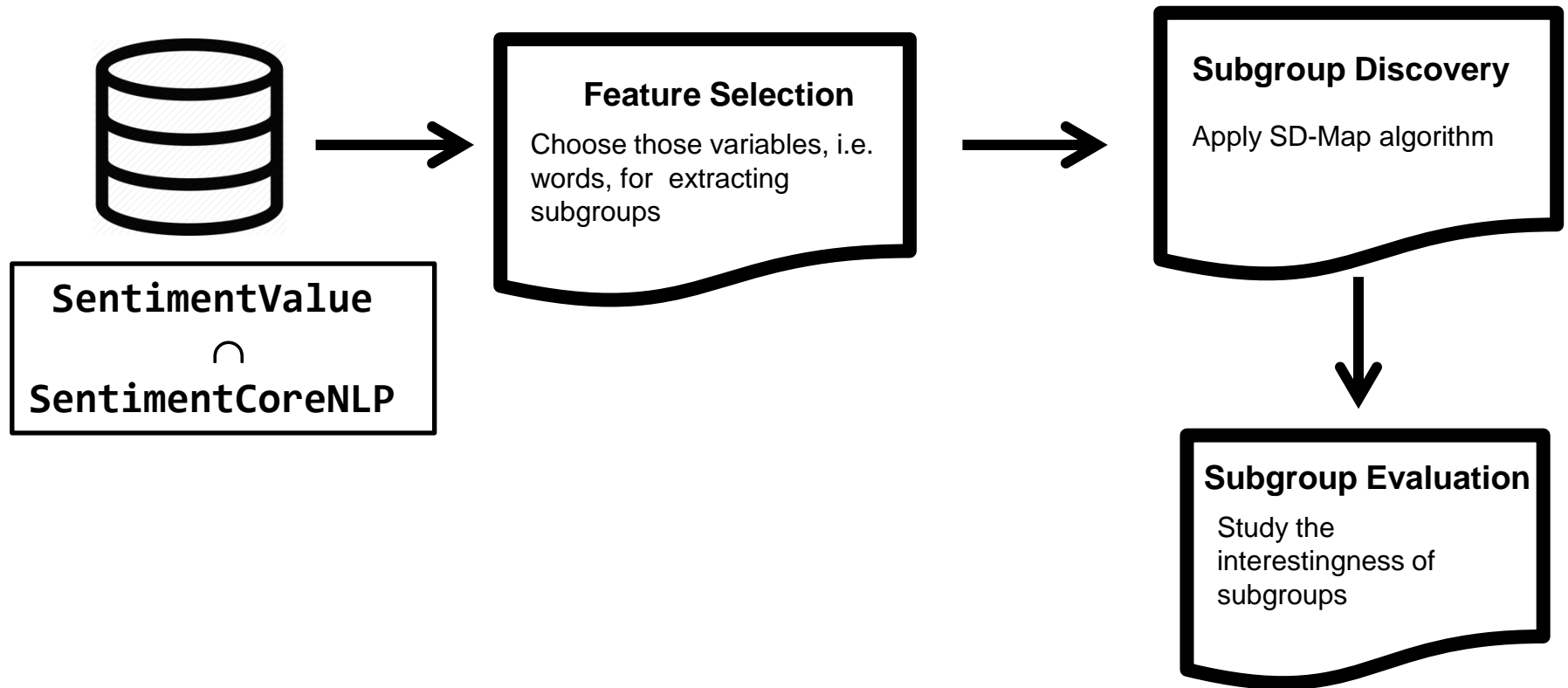
4. SUBGROUP DISCOVERY

INTERESTING PATTERNS



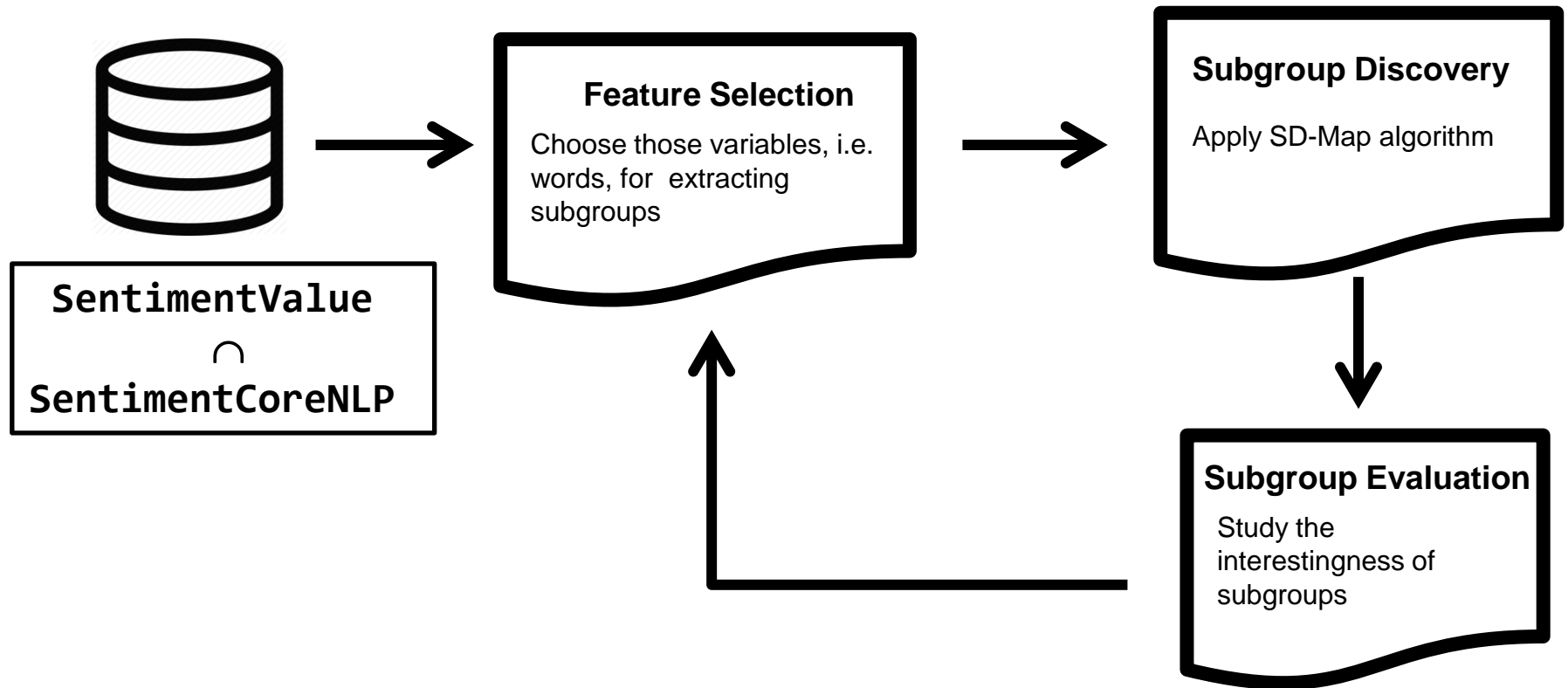
4. SUBGROUP DISCOVERY

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.

5. CONCLUSION

CONCLUSIONS

5. CONCLUSION

CONCLUSIONS

In general...

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

5. CONCLUSION

CONCLUSIONS

In general...

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

5. CONCLUSION

CONCLUSIONS

More precisely...

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

5. CONCLUSION

CONCLUSIONS

More precisely...

1. Suitable TripAdvisor Alhambra **data set** for applying sentiment analysis
2. **Low correlation** between human and machine sentiment

5. CONCLUSION

CONCLUSIONS

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

CONCLUSIONS

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

FURTHER WORK

Next months...

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

5. CONCLUSION

FURTHER WORK

Next months...

1. Talk to **Patronato** about this project
2. Work on the **sentiment labels**

5. CONCLUSION

FURTHER WORK

Next months...

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

5. CONCLUSION

FURTHER WORK

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

FURTHER WORK

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

FURTHER WORK

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!

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

Ana Valdivia García

19 September 2016

Department of Computer Science and Artificial Intelligence



ugr

Universidad
de Granada