Table of Content

1 Introduction **1**

1.1 Hypothesis and Objectives 6

1.2 Methodology 6

1.3 Contribution of this Work 7

1.4 Document Structure 9

2 Preliminaries **10**

2.1 Crisis Management 10

2.2 Data Analysis Tools 11

2.2.1 Classification 12

2.2.2 Support Vector Machine Algorithm 13

2.2.3 Decision Tree 14

2.2.4 Random Forest 15

2.2.5 Neural Networks 15

2.3 Clustering 16

2.3.1 Partitional Models 17

2.3.2 Density Models 19

2.3.3 Evaluation Methods and Metrics 19

2.4 Natural Language Processing (NLP) 21

2.4.1 Word Representation 22

2.4.2 Sequence Labeling 23

2.5 Network Analysis . . . 25
2.5.1 Basic Concepts of Network Analysis . . . 26
2.5.2 Network and Node Descriptives . . . 28
2.5.3 Subgroups and Community Detection Algorithms . . . 29

3 Related Work **31**
3.1 Social Media Messages in Mass Emergency . . . 31
3.1.1 Crisis Characterization . . . 32
3.1.2 Crisis Classification . . . 35
3.1.3 Crisis Detection . . . 37
3.2 Group Polarization in Social Media . . . 39
3.2.1 Content-based Polarization Analysis . . . 40
3.2.2 Network-based Polarization Analysis . . . 40

4 A Domain-Independent Crisis Detection Approach **44**
4.1 Proposed Approach . . . 45
4.1.1 Dataset Description . . . 46
4.1.2 Data Pre-Processing . . . 47
4.1.3 Signal Creation . . . 48
4.1.4 Time-Window . . . 51
4.1.5 Geographic Spread . . . 52
4.2 Methodology . . . 53
4.2.1 Independent Analysis of Hierarchies . . . 55
4.2.2 Dependent Analysis of Hierarchies . . . 56
4.2.3 Geographic Spread Analysis . . . 57
4.2.4 Crisis Detections in the Wild . . . 58
4.3 Discussion . . . 60

5 Cross-Lingual and Cross-Domain Crisis Message Classification **62**

5.1 Proposed Approach . . . 63
5.1.1 Transfer-Learning Scenarios . . . 63
5.1.2 Data Representations . . . 65
5.2 Unified Dataset Construction . . . 66
5.2.1 Label Merging . . . 67
5.2.2 Crisis Categorization . . . 67
5.2.3 Unified Dataset Description . . . 68
5.3 Experimental Setup . . . 68
5.3.1 Model Training . . . 69
5.3.2 Model Testing . . . 70
5.4 Results . . . 71
5.4.1 English Classification . . . 71
5.4.2 Spanish Classification . . . 73
5.4.3 Italian Classification . . . 74
5.4.4 Results by classification scenario . . . 75
5.5 Discussion . . . 76

6 A Comparative Study of Communication Patterns Across Crisis Events 78

6.1 Crisis Event Dataset Creation . . . 79
6.1.1 Dataset Description . . . 79
6.1.2 Annotation of Crisis Dimensions . . . 80
6.2 Feature Extraction . . . 81
6.3 Crisis Analysis . . . 83
6.3.1 Crisis Similarity . . . 84
6.3.2 Crisis Evaluation . . . 85
6.4 Discussion . . . 88

7 Discussion Analysis during Crises 93

7.1 The 2019 Chilean Social Unrest Movement 94
7.2 Dataset 95
7.3 Proposed Approach 96
7.3.1 Detecting communities 96
7.3.2 Framing assignment 98
7.3.3 Content analysis among communities 100
7.4 Discussion 104

8 Conclusions and Future Directions **108**
8.1 Conclusions 108
8.2 Evaluation of the Thesis Objectives 109
8.3 Future Directions 111

Bibliography **129**