

Table of Contents

Preface	v
Chapter 1: Fundamentals of Mining	1
Social media and its importance	1
Various social media platforms	3
Social media mining	4
Challenges for social media mining	4
Social media mining techniques	6
Graph mining	6
Text mining	7
The generic process of social media mining	7
Getting authentication from the social website – OAuth 2.0	8
Differences between OAuth and OAuth 2.0	10
Data visualization R packages	10
The simple word cloud	11
Sentiment analysis Wordcloud	12
Preprocessing and cleaning in R	14
Data modeling – the application of mining algorithms	14
Opinion mining (sentiment analysis)	14
Steps for sentiment analysis	15
Community detection via clustering	18
Result visualization	19
An example of social media mining	19
Summary	20
Chapter 2: Mining Opinions, Exploring Trends, and More with Twitter	21
Twitter and its importance	21
Understanding Twitter's APIs	23
Twitter vocabulary	23

Creating a Twitter API connection	24
Creating a new app	25
Finding trending topics	28
Searching tweets	29
Twitter sentiment analysis	30
Collecting tweets as a corpus	30
Cleaning the corpus	32
Estimating sentiment (A)	35
Estimating sentiment (B)	39
Summary	54
Chapter 3: Find Friends on Facebook	55
Creating an app on the Facebook platform	56
Rfacebook package installation and authentication	58
Installation	58
A closer look at how the package works	59
A basic analysis of your network	62
Network analysis and visualization	64
Social network analysis	64
Degree	66
Betweenness	67
Closeness	68
Cluster	68
Communities	69
Getting Facebook page data	71
Trending topics	73
Trend analysis	73
Influencers	74
Based on a single post	74
Based on multiple posts	76
Measuring CTR performance for a page	77
Spam detection	80
Implementing a spam detection algorithm	80
The order of stories on a user's home page	84
Recommendations to friends	87
Reading the output	89
Other business cases	90
Summary	90

Chapter 4: Finding Popular Photos on Instagram	93
Creating an app on the Instagram platform	94
Installation and authentication of the instaR package	96
Accessing data from R	97
Searching public media for a specific hashtag	97
Searching public media from a specific location	98
Extracting public media of a user	99
Extracting user profile	99
Getting followers	100
Who does the user follow?	101
Getting comments	102
Number of times hashtag is used	104
Building a dataset	105
User profile	106
User media	107
Travel-related media	108
Who do they follow?	109
Popular personalities	110
Who has the most followers?	110
Who follows more people?	111
Who shared most media?	111
Overall top users	112
Most viral media	112
Finding the most popular destination	113
Locations	114
Locations with most likes	115
Locations most talked about	115
What are people saying about these locations?	116
Most repeating locations	117
Clustering the pictures	118
Recommendations to the users	123
How to do it	123
Top three recommendations	130
Improvements to the recommendation system	131
Business case	132
Reference	132
Summary	133

Chapter 5: Let's Build Software with GitHub	135
Creating an app on GitHub	136
GitHub package installation and authentication	139
Accessing GitHub data from R	141
Building a heterogeneous dataset using the most active users	142
Data processing	144
Building additional metrics	145
Exploratory data analysis	148
EDA – graphical analysis	150
Which language is most popular among the active GitHub users?	150
What is the distribution of watchers, forks, and issues in GitHub?	153
How many repositories had issues?	156
What is the trend on updating repositories?	157
Compare users through heat map	158
EDA – correlation analysis	161
How Watchers is related to Forks	162
Correlation with regression line	163
Correlation with local regression curve	164
Correlation on segmented data	165
Correlation between the languages that user's use to code	166
How to get the trend of correlation?	168
Reference	171
Business cases	172
Summary	173
Chapter 6: More Social Media Websites	175
Searching on social media	176
Accessing product reviews from sites	180
Retrieving data from Wikipedia	181
Using the Tumblr API	190
Accessing data from Quora	196
Mapping solutions using Google Maps	198
Professional network data from LinkedIn	203
Getting Blogger data	208
Retrieving venue data from Foursquare	211
Use cases	218
Yelp and other networks	218
Limitations	219
Summary	219
Index	221