

Project Outline

Data Mining Team Project

presented by
Sophia Mai
Ivan Karachunskiy
Timo Strauch
Natalie Buchner submitted to the

Data and Web Science Group
Prof. Dr. Bizer
University of Mannheim

April 2016

Contents

1	Project Proposal	1
1.1	Our problem	1
1.2	The Yelp Dataset	2
1.2.1	Background	2
1.2.2	Integration of JSON	2
1.3	Our approach	2
1.3.1	Preprocessing	2
1.3.2	Algorithms	2
1.4	Measure of Success	2
1.5	Possible Results	2

Chapter 1

Project Proposal

1.1 Our problem

When people want to choose which restaurant, bar, cafe etc to go they usually take a look at specific properties of the business, such as location, price range, opening hours, cuisine and user based rating. As a result these properties play a decisive role in how successful businesses are. There is a vast amount of businesses which are quite similar if you only take those rather general properties into account, but still their economical success often differs significantly. As a consequence there must be other determining factors that make people choose one business over another.

Some of these factors can be found when taking a look at textual reviews where people state their opinion on certain aspects of the business and often give a concluding rating (e.g. through stars at a scale of 1 to 5). The goal of this project is to extract new aspects of the business from reviews that can be used to give a more sophisticated and granular view on the restaurant or bar.

1.2 The Yelp Dataset

1.2.1 Background

1.2.2 Integration of JSON

1.3 Our approach

1.3.1 Preprocessing

1.3.2 Algorithms

1.4 Measure of Success

1.5 Possible Results