UNIVERSITY OF SOUTHERN DENMARK

SOFTWARE ENGINEERING 6. SEMESTER

Datamining and its use

Author:
Lasse Bjørn HANSEN
Simon FLENSTED

Supervisor:
Jan Corfixen Sørensen SØRENSEN

A report submitted in fulfillment of the requirements
of Software Engineering 6. semester

at

University of Southern Denmark
TEK

February 7, 2017

"Some quote"

- Gruppe 3

Contents

Chapter 1

Problem statement

1.1 Problem description

The initial problem/challenge is given to us by the company Struct A/S and is described as follows:

When launching sites, whether it being regular websites or web shops, a lot of user activity is logged. We therefore have a large amount of data associated with each of our sites but do not currently use it.
In the future we would like to be able to use logged data to generate an insight into the user activity on our site and actively use this data to create a personalized experience for the users.
The personalized experience can be in the form of product recommendations based on the users activity on the site.

1.2 Problem statement

The data we have been given is in a de-normalized format and the problem therefore comes with two challenges - normalizing the data and utilizing the data to create a personalized experience for the users.
This leads to the following research questions:

- How can you effectively normalize large amounts of data?
- How can you optimally store and access data in a scalable way?
- How can you effectively analyze large amount of data and draw conclusions from it?