



User Behavior Analysis Using Decision Trees

Martin Nicklas Jørgensen

Department of Computer Science



Company Project

Martin Nicklas Jørgensen

Disposition

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- Introduction
- Problem Analysis
- Results & Conclusion
- Questions



Company Project

Martin Nicklas Jørgensen

Disposition

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- Introduction
- Problem Analysis
- Results & Conclusion
- Questions



Company Project

Martin Nicklas Jørgensen

Disposition

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- Introduction
- Problem Analysis
- Results & Conclusion
- Questions



Company Project

Martin Nicklas Jørgensen

Disposition

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- Introduction
- Problem Analysis
- Results & Conclusion
- Questions



Company Project

Martin Nicklas Jørgensen

Simplesite ApS

Introduction

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- Hosts and sells a website CMS.
- Founded in 2001 as *Elk Consulting ApS*, later *123hjemmeside ApS* and now *Simplesite ApS*.
- 20-49 employees globally.
- 400.000 new websites every month.
- 80.000 paying subscribers.



Company Project

Martin Nicklas Jørgensen

Problem Description

Introduction

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- After the change to freemium, a huge increase in new free users was observed.
- The number of paying users did not increase comparatively.
- Many users stop being active after a few days.
- Using Data Science, is it possible to find a pattern in how users who stay are using the site?



Company Project

Martin Nicklas Jørgensen

CRISP-DM

Problem Analysis

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

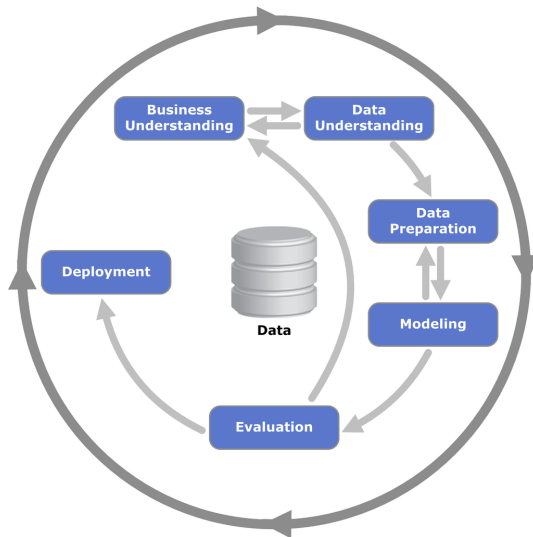


Figure : Diagram of the CRISP-DM method. Image source: Wikimedia Foundation.

Company Project

Martin Nicklas Jørgensen

Business Understanding

Problem Analysis

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- Some users do not stay active for very long, even on the free product.
- Can we figure out what makes users stay?



Company Project

Martin Nicklas Jørgensen

Data Understanding & Preparation

Problem Analysis

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- EngagementData **datasets**.
- CustomerJourney **datasets**.
- Features are removed from the datasets if they are derivative or not relevant.
- Datasets are merged into a single dataset.
- Data from Sep. 2015 are used for training. (463716 observations)
- Data from Oct. 2015 are used for test. (495390 observations)
- Final dataset have 15 features one of which is the target variable *iscjretained*.



Company Project

Martin Nicklas Jørgensen

Modelling - Tree Type

Problem Analysis

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

Max Depth	<code>rpart</code> Accuracy	<code>ctree</code> Accuracy
4	94.2799 %	94.27990 %
8	94.2799 %	94.31958 %
12	94.2799 %	94.36638 %

Table : The mean accuracy for the different 5-fold cross validation runs.



Company Project

Martin Nicklas Jørgensen

Modelling - Formula & Depth

Problem Analysis

Formula	Max Depth	Mean Accuracy
iscjretained ~ .	4	94.27990 %
	6	94.29672 %
	8	94.31958 %
iscjretained ~ edits14	4	93.50055 %
	6	93.50227 %
	8	93.50119 %
iscjretained ~ logins14	4	94.27990 %
	6	94.27990 %
	8	94.27990 %
iscjretained ~ edits14 + logins14	4	94.27990 %
	6	94.28465 %
	8	94.29414 %

Table : Mean accuracy of different formulas and tree depths using 5-fold cross



Company Project

Martin Nicklas Jørgensen

Evaluation - Dataset Bias

Problem Analysis

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

Dataset	TRUE	FALSE
<i>Training</i>	30358	433358
<i>Test</i>	40731	454659
<i>Equal</i>	30358	30358

Table : The distribution of the *iscjretained* target variable classes in the different datasets.



Company Project

Martin Nicklas Jørgensen

Deployment

Problem Analysis

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

- Deployment was not done during this project.
- Mail 2.0
- Possible design mentioned Future Work.



Company Project

Martin Nicklas Jørgensen

Results

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

Maximum Depth	Accuracy
4	92.84039 %
6	92.84846 %
8	92.75823 %

Table : The results of the final datarun when training on the full training set and trying to predict the entire test set.



Company Project

Martin Nicklas Jørgensen

Results

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

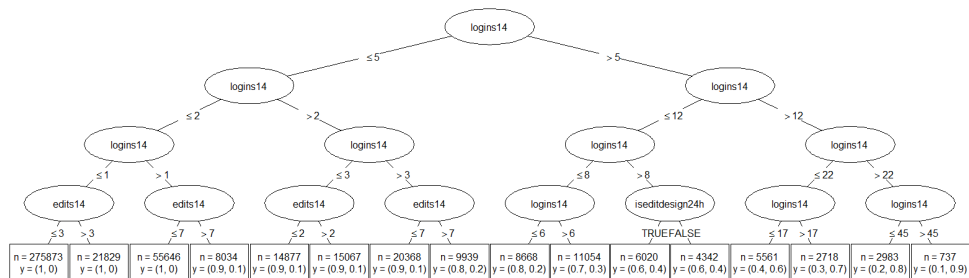


Figure : The conditional inference tree produced by the code when using a maximum depth of 4.



Company Project

Martin Nicklas Jørgensen

Results

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

Maximum Depth	Accuracy
4	91.79051 %
6	91.74852 %
8	91.74388 %

Table : The results of the final data run when training on the full training set excluding the *logins14 variable* and trying to predict the entire test set.



Company Project

Martin Nicklas Jørgensen

Results

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

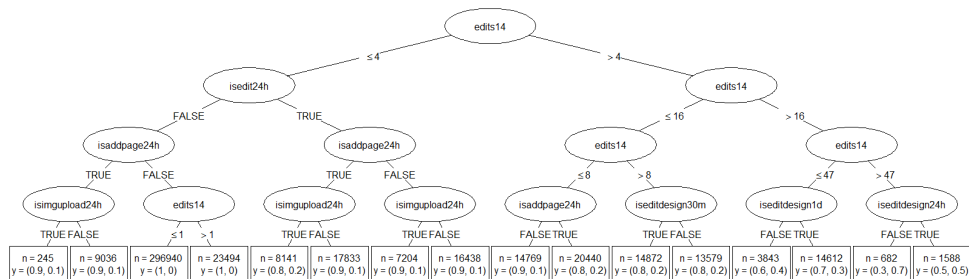


Figure : The conditional inference tree produced by the code when using a maximum depth of 4 and excluding the *logins14* attribute.



Company Project

Martin Nicklas Jørgensen

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

Future Work

Conclusion and Future Work

- Better datasets.
- Rough design of automated solution.



Company Project

Martin Nicklas Jørgensen

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

Conclusion

Conclusion and Future Work

- TODO: Conclude things in here as well.



Company Project

Martin Nicklas Jørgensen

Disposition

Introduction

Problem Analysis

Results

Conclusion and Future Work

Question

Questions

- Questions.

