

Business needs of the Final Paper

for the CEU MSc in Business Analytics program

Gabor Horvath

1 Introduction

The structure of the document follows the Cross Industry Standard Process for Data Mining (CRISP-DM) process model, which is a non-proprietary, documented, and freely available data mining model (Shearer 2000). Whenever the model sections can be matched to (and can fulfill) the requirements stated by CEU for the Final Paper I'm using the appropriate section identified by the CRIPS-DM model. Please keep in mind that the model supports the full end-to-end process of a data mining project, but the project does not require the use of all the model elements.

2 Business understanding

2.1 Determine Business Objectives

2.1.1 Business Objectives

TODO

2.1.2 Business Success Criteria

- Identification of featured determining the risk potential of an airport
- Working model for animal strike prediction

2.2 Assess Situation

2.2.1 Inventory of resources

- Flight Data
- Animal Strike Data
- R
- Buckets

2.2.2 Requirements, Assumptions, and Constraints

2.2.3 Risks and Contingencies

- No initial Risks identified
- No initial Contingencies identified

2.2.4 Terminology

TODO

2.2.5 Costs and Benefits

This is a one-man project, no significant cost is expected. Main benefit is to put to and almost end-to-end scenario the topics covered during the courses and discovering bits and bolts of the techniques for creating the project.

2.3 Determine Data Mining Goals

2.3.1 Data Mining Goals

- Understand, Analyse, Clean and Merge the source data correctly
- Create the required attributes
- Generate the required records (if applicable)

2.3.2 Data Mining Success Criteria

- Identification of featured determining the risk potential of an airport
- Working model for animal strike prediction

2.4 Produce Project Plan

2.4.1 Project Plan

The project is managed in an agile way, where all the tasks, requirements, issues, solutions, and ideas are kept in a project @ Buckets.

2.4.2 Initial Assessment of Tools and Techniques

- Programming language:
 - R
- GUI for the programming language:
 - RStudio
- Documentation (including this initial project description) is created using:
 - knitr
 - MiKTeX
 - ReporteRs
- Data visualisation:
 - ggplot2
- Data manipulation:
 - access2csv
 - dtplyr
- Project plan / task management:
 - Buckets
- Source code repository:
 - GitHub

3 Data Understanding

3.1 Collect Initial Data

3.1.1 Initial Data Collection Report

This report will be part of the following documents:

- Preliminary Report
- Final Paper

3.2 Describe Data

3.2.1 Data Description Report

TODO

3.3 Explore Data

3.3.1 Data Exploration Report

This report will be part of the following documents:

- Preliminary Report
- Final Paper

3.4 Verify Data Quality

3.4.1 Data Quality Report

This report will be part of the following documents:

- Preliminary Report
- Final Paper

4 Contributors

Student: Gabor Horvath

Mentor: Gergely Daroczi

References

Shearer, Colin. 2000. "The CRISP-DM Model - The New Blueprint for Data Mining" 5 (September). Journal of Data Warehousing: 13–22.