

9/18/2015

TEAM PLUS

Nick D. wondergem, upadhyay ishant, kichukov veselin, velev dan, draganov dragan, ngabo francky

Minor report

ORIKAMI

Document history

Revisions

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Status | Date | Changes |
| 0.1 | concept | 18-09-2015 | Template created and introduction |
| 0.2 | concept | 20-09-2015 | Assignment formulated and info Team Orikami |
| 0.3 | concept | 22-09-2015 | Planning and general additions/changes |
| 0.4 | concept | 13-09-2015 | Background and approach |
| 1.0 | draft | 27-09-2015 |  |

Approval

This document needs to be approved by:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Approval date | Name | Function | Paraph |
| 1.0 |  | Theo Cats | Tutor |  |
| 1.0 |  | Bram den Teuling | Product owner |  |

Distribution

This document has been sent to:

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Send date | Name | Function |
| 1.0 | 28-09-2015 | Theo Cats | Tutor |
| 1.0 | 28-09-2015 | Bram den Teuling | Product owner |

Contents

[Introduction 2](#_Toc441582118)

[Team orikami 4](#_Toc441582119)

[Background 5](#_Toc441582120)

[Assignment 7](#_Toc441582121)

[Approach 9](#_Toc441582122)

[Planning 10](#_Toc441582123)

[Agreements about communication 11](#_Toc441582124)

[Additional information 11](#_Toc441582125)

[Appendix A: Gant chart 12](#_Toc441582126)

# Introduction

The growth of data – both structured and unstructured – will present challenges as well as opportunities. Team Plus will get the opportunity to work together with orikami. The proposed challenge from orikami is to classify the different stages in the lifecycle of a song.

For this project we will be using the scrum process, some key points:

* product owner: orikami
* development team: Team Plus
* scrum master: a different member of the development team per sprint
* sprint planning:
* sprint duration: 3 weeks
* sprint review: feedback on deliverables after each sprint
* backlog management: chrome app Kerika

Team Plus is thrilled to work on this project, because 'Big Data' captures what makes code exciting in the first place: doing massively valuable macro-scale things simply by manipulating symbols according to formal rules. During this exciting time we will aim to deliver:

* project document
* exploratory plots
* visualizations
* 'life-cycle' templates

# Team orikami

****

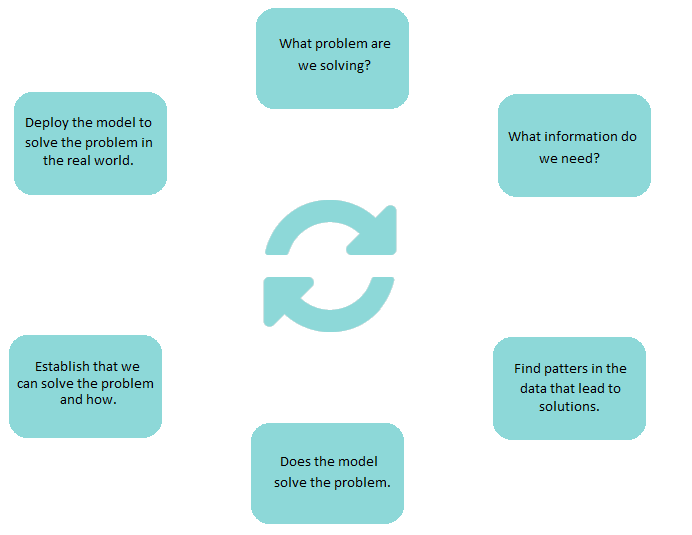
# Background

Location

Orikami is a data science company which is specialized in *BIG DATA.* It is located in Nijmegen in the west of the Netherlands. It started in 2013 and is composed by nine members.

At Orikami the collaborate, partner-up and co-create.

Agile workflow



Partners

It has many partners such as *D&B*, *Radboudumc, casemix, Ministerie van Onderwijs cultuur en wetenschap, Inspecties van Onderwijs cultuur en wetenschap … etc.*

Products

Orikami operates in many fields by turning *Data* into products (*Connecting mobility, YouBeDo, IRMA, BierApp, U-map,* *Mozart…etc.)*

*** Connecting mobility*:** Strives to integrate all initiatives in the field of Intelligent Transport System (ITS) in the Netherlands.

*** YouBeDo:*** Their online bookstore offers two million books for sale. Orikami help them search this vast amount of book titles. Orikami developed the search engine behind YouBeDo.com and are analyzing its performance on conversion.

*** IRMA:*** Integrity Risk Management App, is created for organizations to help them assess their integrity risks.

*** U-map:*** *The* multi-dimensional classification is gathered from universities. Source information is rigorously checked and standardized over countries and measurement standards. The mapping gives rise to unforeseen connections and similarities between universities.

*** BierApp:*** It gives you advice on what to drink next and shows the beers listings of cafes. The ‘BierApp’ makes sure you always drink the right beer on the right occasion.

*** Mozart:*** Mozart listens to open sources like Twitter, Facebook, YouTube and Last.FM to learn what people are saying. We combine this with specialized information from radio airplay, Spotify and YouTube. The last piece of information comes from a large user panel. We sample feedback from users about what they like.

Etc …

They deliver a real time insights, self-learning algorithms, and try to accurate prediction.

# Assignment

Description

With the data provided from Orikami, such as tweet feeds and surveys of songs, we have to apply our knowledge of data science to present the life cycle of songs. We will analyze the data, in order to be able to visualize and classify different lifecycles later on. Hence, this will contribute to the Orikami’s ‘*Mozart’* project, which basically learns what think about music through social media.

Purpose of the assignment

Orikami would like to improve its existing project, *‘Mozart’,* with providing a classification of the different stages in the lifecycle of a song. Therefore, this will help the user to get a clear view of when people are tired of a song, the peak time, the investment, profit of it and etc.

The assignment would provide an insight into the ‘sweet spot’ for marketing a particular song. We will give our best to create a template for the classification of different lifecycles.

Analysis of the assignment

*The main question in our assignment is how to present the life cycle of the songs.*

*The problem that is needed to be solved is trying to understand the tendency for music from the data provided from Orikami.*

*The challenge of the assignment is “diving in the ocean of tweets in search for life cycle of the songs”.  From given tweet feeds and information about the songs we have to be able to accomplish our research.*

*The problem exists because of the enormous amount of data and different people’s taste of music.*

*It is urgent because of the need for better understanding the lifecycle of the songs.*

*The consequence if nothing is done about it is misunderstanding about the lifecycle of the songs.*

Activities

**IN Scope:**

* Cleaning data **-** we have to make sure that the quality of the data is good, hence we will apply few algorithms to improve the quality
* Analyze data **-** we will explore and analyze the data, provided by Orikami, to find insights in it.
* Look for outliers **-** to find out the co-relations between the data.
* Apply ML algorithm **-** once the co-relations are revealed we can apply a ML algorithm which will define by itself the lifecycle of a song.
* Showing the outcome **-** visualizing the end result and showing it to the client

**OUT Scope:**

* Knowledge not yet gained **-** since we are still students, we will need to learn and solve unknown, to us, errors and problems by ourselves.
* Answering questions that are important for us **-** it is possible to come up with questions regarding our work.
* Translating messages/feeds **-** the provided data is from all around the world and mostly, Netherlands, hence we will need to translate some of the data.
* Data size **-** data may be too big to be stored on our PCs, therefore we will need to use Fontys’ server to store it.

**Risks:**

* Different languages **-** hence we might get a wrong understanding of the data while translating
* RFC (Request For Change) **-** we might encounter unforeseen issues that take a lot of time to study them, hence a deliverable might not be delivered.
* Urban phrases**-** phrases that are difficult for us to understand and translate.
* End result not clear **-** end result might be not very clear to the client, hence we need to explain it as simple as possible.

**Must, Should, Could, Won’t**

|  |  |  |  |
| --- | --- | --- | --- |
| **Must** | **Should** | **Could** | **Won’t** |
| 1.Cleaning data | 1.Apply  ML algorithm |  | Implementation into Orikami’s project |
| 2. Analyze data |  |  |  |
| 3.  Look for outliers |  |  |  |
| 4.  Showing the outcome |  |  |  |

# Approach

Description

We as a team believe that using the scrum approach is best in this situation. It allows for flexibility which will help with reaching our goals faster allowing changes to be easily made in case of mistakes and overall a better team dynamic.

Methods

We will be using sprints which consist of planning what needs to be done, getting it approved, splitting the work into parts and finally working. We will have 3 meetings a week to briefly discuss what everyone has done to contribute, did they meet any obstacles, and how are we going to resolve the issues at hand. In addition we will keep a backlog of everything to see if a similar problem occurred in the past so we can resolve it as quickly as possible.

Map

Step 1. Exploration

At first we will begin by cleaning the data and selecting specific sets to achieve an optimal number of variables. With those things in mind we will be able to produce visual representation of the data.

Step 2. Model building and validation

We will search and find a model that best suites our project. The way we are going to do that is by testing the models on a select dataset and see which one produces the best predictive results.

Step 3.  Deployment

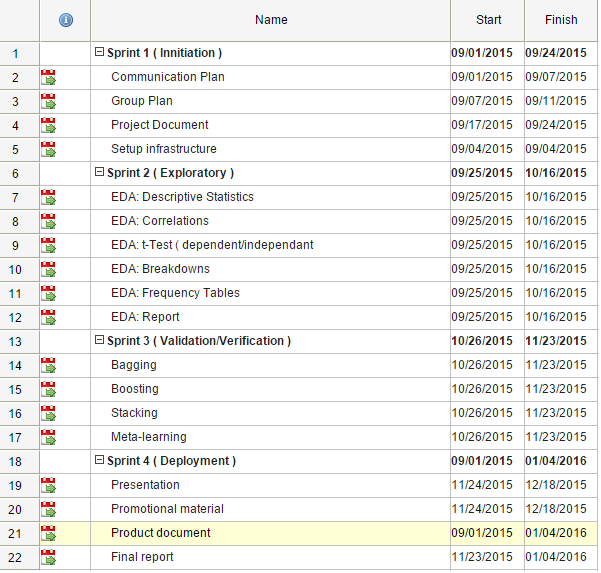
That final stage involves using the model selected as best in the previous stage and applying it to new data in order to generate template helping visualizing the dataset.

Final product

The final product will be a template that gives information about the life cycle of songs. With it you will be able to visually see the life cycle of a song of your choosing. There will be clusters of people shown based on how did they react to the song, if they liked it or didn’t, where they live (maybe a song is more popular in Eindhoven than in Amsterdam for instance) and more. The characteristics of the song will be shown. For instance duration, beat, genre and more.

These features will help you to make a more informed and optimal marketing strategy to maximize profit figures in the foreseeable future.

# Planning



\*Gant chart in Appendix A

# Agreements about communication

We have agreed to use all possible means of communication available including meeting, mail, social media (Facebook) etc.

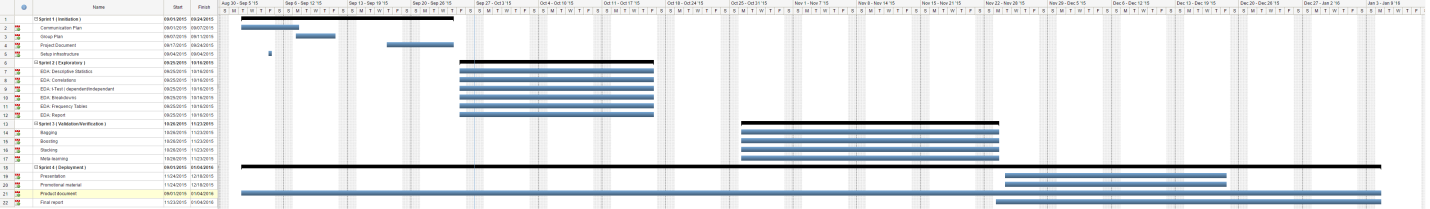
|  |  |  |  |
| --- | --- | --- | --- |
| Whom | Purpose | Frequency | How |
| Saiden(Orikami) (technical support) | About Detail information about Project | 2 times | Mail |
| Theo Cats (tutor) | About how to start and process to follow for Project | 1 | Meeting |
| Team+ Members | About division of work, total work in progress | 3 times | Meeting ,mail, social media(Facebook) |

# Additional information

Here are the things which may be affect the Project and the resolution for the reason:

* Since we are student so we may not be perfect at our work but we will learn the required techniques needed for work and complete the work.
* If for any reason some of our group member fall sick or may be absent due to some personal reason and if there is any delay in our project, we will redesign the division of work among the rest of the members and try to finish it in time.

# Appendix A: Gant chart



full version see appendix\_a\_gant\_chart.bmp