

University College of Northern Denmark

Technology and Business

Computer Science Academy Profession (AP) Degree

Class: DMAJ0916 (Big Data)

Title: WASP Report– 4th Semester 2018

Project participants (Group 2):

-Ralfs Zangis

-Andrei-Eugen Birta

Supervisors:

-Nadeem Iftikhar

Due date: 2018-May-24 (14:00 UTC+1)

Submission date: 2018-December-17

Birta Zanigis

Table of Contents

[1. Introduction 2](#_Toc512783065)

[2. Problem Statement 2](#_Toc512783066)

[a. Introduction 2](#_Toc512783067)

[b. Case Description 2](#_Toc512783068)

[c. Learning Goals 3](#_Toc512783069)

[3. Development Framework 3](#_Toc512783070)

[a. Pros 4](#_Toc512783071)

[b. Cons 4](#_Toc512783072)

[4. Development Process 4](#_Toc512783073)

[a. Data Acquisition 4](#_Toc512783074)

[b. Data Wrangling 4](#_Toc512783075)

[c. Descriptive Analysis 4](#_Toc512783076)

[d. Diagnostic Analysis 4](#_Toc512783077)

[e. Predictive Analysis 4](#_Toc512783078)

[5. Conclusion 4](#_Toc512783079)

[a. Denouement 4](#_Toc512783080)

[b. References 4](#_Toc512783081)

# Introduction

This document summarizes the collaboration of Group 2 for the Big Data specialization exam of the 4th Semester. The group consists of 2 members of 2 different nationalities. Despite the major differences in our opinions, we agreed on certain rules and guidelines to follow, thoroughly elaborated in the accompanying document called group contract; we managed to harness the benefits of diverse ideas and identify multiple possible approaches to certain problems, be it project or working regulations.

# Problem Statement

## Introduction

The purpose of this project is creating a machine learning algorithm that is capable of predicting to a higher degree of accuracy, next possible collisions between aircrafts and wildlife animals on a worldwide scheme, considering both civil and military aircrafts.

The purpose of this project is applying various Big Data analyzing techniques on a national dataset regarding collisions between, both civil and military aircrafts, and wildlife animals.

## Case Description

Day to day activities thought us that, the collision between something massive and something small and frail, usually ends up pretty bad for the small object and affects little to not at all the massive object, when it comes to aircraft collisions with birds and other wildlife creatures, things tend to go bad for both parties. Usually killing the animal and ruining the aircraft, possible for the rest of its “life”.

The following are images of possible damage that such a collision can cause, to an aircraft.

And considering that wildlife population is fluctuating depending on different seasons of the year, but usually increasing in numbers, such collisions should be taken with all seriousness and evaded as much as possible.

## Learning Goals

Some of the learning goals for this project are:

* Gathering useful datasets related to the case (Data acquisition)
* Converting datasets to a common format, in order to facilitate data analysis, using tools offered by Python
* Filtering data (dealing with missing values, misspelled words or wrong datatypes), using tools offered by Python
* Describing what and why it has happened, using descriptive and diagnostic analysis techniques
* Predicting possible collision areas, at different times of the year and possibly wildlife species and types of damage, to facilitate warning emissions by concerned authorities

# Development Framework

Following previously acquired knowledge, from the 3rd Semester’s System Development course, we decided that the best way of choosing a development method is by evaluating the team and creating a Boehm and Turner Model.

The following image is the diagram we have come up with, following the self-evaluation process.



From the diagram above, resulted that we needed some kind of agile development method, due to the high amount of expected changes, small team size, and team’s culture, but is structured enough to accommodate for the project’s criticality.

Although the high level of criticality indicates that structured development method should be used, we have decided to work following the Kanban development method, because of two reasons: it best fits the other four measurements and it’s quite unlikely that our project will be used by any company, due to various reasons.

## Pros

* The small number of guidelines gives the programmer the freedom to work as he pleases
* The Kanban board helps manage the project responsibilities and keep track of what has been achieved
* The Kanban board allows for a better understanding of work and workflow.

## Cons

* Lack of “urgency” concept in the Kanban board, can lead to unnecessary waste of time, due to dependencies of certain tasks on other tasks

# Development Process

Typically the development process for a Big Data project starts from one or more small and clearly defined questions, followed by Data Acquisition, Data Wrangling, Descriptive Analysis, Diagnostic Analysis, Predictive Analysis and ending with Perspective Analysis; all of them bringing important additions to the overall meaning of the project and helping those who are concerned about the matter, better understand the situation and take actions based on facts not on feelings.

## Data Acquisition

Data Acquisition is the first step that has to be made when working on a Big Data related project. This step refers to acquiring the necessary data for answering the previously defined questions.

Our datasets were acquired from trusted websites that hold thousands and thousands of various datasets, the exact links for those datasets can be seen in the “References” part of this report.

## Data Wrangling

## Descriptive Analysis

## Diagnostic Analysis

## Predictive Analysis

# Conclusion

## Denouement

## References

* Diagram taken from “Balancing Agile with Discipline” by Barry Boehm Richard Turner
* https://wildlife.faa.gov/databaseSearch.aspx - the original dataset
* https://www.kaggle.com/faa/wildlife-strikes - dataset from kaggle (could be used as example)
* https://www.transportation.gov/ - for more info about transport
* https://wildlife.faa.gov/downloads/StrikeReport1990-2012.pdf - info about bird strikes (could be used in report and presentation)
* http://aircharterguide.com/Operators (filling blank info)