## Assignment Details

Module Code: **ICS5110**

Module Name: **Applied Machine Learning**

Assignment Weighting: 80%

Report Authors

* **Gabriel Sammut (184395M)**
* **Nicholas Frendo**

Contents

[Assignment Details 1](#_Toc496803863)

[Section 1 – Introduction 2](#_Toc496803864)

[Chosen Dataset 2](#_Toc496803865)

[Machine Learning Techniques 3](#_Toc496803866)

[Section 2 – Background 3](#_Toc496803867)

[Section 3 – Experiments 3](#_Toc496803868)

[Section 4 – Conclusions 3](#_Toc496803869)

[References 3](#_Toc496803870)

# Section 1 – Introduction

This practical assignment looks into the utilisation of four machine learning techniques on a self-chosen data set. [1] However this report will specifically focus on the application of two of the implemented techniques. The objective of this study is to compare and monitor a selected number of machine learning algorithms by applying them to the afore mentioned dataset and reporting in detail the observed results. The report will feature an apt comparison between the performance and applicability of the machine learning algorithms as applied to the chosen dataset, as well as detailing each of the methods in a manner of detail. Furthermore the implemented algorithms were not picked under any particular bias, and merely implemented to detail the capabilities, distinction and applicability when applied to different datasets.

## Chosen Dataset

The tragedy known popularly as the sinking of the RMS Titanic is a very popular event. On April 15th of 1912, the Titanic sank after colliding with an iceberg, during the ship’s maiden voyage. Out of the 2224 aboard the vessel, a combination of 1502 crew and passengers were killed during this disastrous accident. Amongst the many reasons that resulted in the high fatality rate in the wake of the accident, was the lack of lifeboats available for the people on board. Partially considered an element of luck, some groups of people were more likely to survive than others, including but not limited to the women and children on board the ship, as well as those from the upper-class. [1]

Using this particular dataset allows us to analyse what sort of people were likely to survive the calamitous event. Composed of twelve data fields, the dataset consists of a total of 1309 candidates. Each candidate is summarised using the following descriptors:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Definition** | **Key** |
| PassengerId | Passenger Identification |  |
| Survived | Survival | 0 = No, 1 = Yes |
| Pclass | Ticket class | 1 = 1st, 2 = 2nd, 3 = 3rd |
| Sex | Sex | 0 = Male, 1 = Female |
| Age | Age in years |  |
| Sibsp | # of siblings / spouses aboard the Titanic |  |
| Parch | # of parents / children aboard the Titanic |  |
| Ticket | Ticket number |  |
| Fare | Passenger fare (ticket cost) |  |
| Cabin | Cabin Number |  |
| Embarked | Port of embarkation | C = Cherbourg,  Q = Queenstown,  S = Southampton |

It is worth noting the following assumptions which were considered when applying our analysis on this particular dataset:

* **Pclass**: A proxy for socio-economic status (SES).
* **Age**: Age is a fractional value when less than one.
* **Sibsp**: The dataset defines family relationships as:
  + *Sibling –* brother, sister, stepbrother, stepsister.
  + *Spouse* – Husband, wife (Mistresses and fiancés are ignored).
* **Parch**: The dataset defines family relations as:
  + *Parent* – mother, father
  + *Child* – daughter, son, stepdaughter, stepson

## Machine Learning Techniques

A mixture of supervised and unsupervised techniques were utilised, totalling up to four machine learning algorithms being applied to the Titanic dataset:

* Naïve Bayes Classification
* Support Vector Machine Classification

# Section 2 – Background

The following consists of a detailed explanation regarding the inner workings of the applied machine learning algorithms.

# Section 3 – Experiments

# Section 4 – Conclusions

# References

[1]"Titanic: Machine Learning from Disaster | Kaggle", Kaggle.com, 2017. [Online]. Available: https://www.kaggle.com/c/titanic/data. [Accessed: 26- Oct- 2017].