

Azure PipeLine / Shiny app :   
Credit Card Fraud Modeling and Deployment

**Proposal**

Word Count: 1000

Higher Diploma in Science in Data Analytics

10524150 / Ciaran Finnegan

10524150@mydbs.ie / ciaran@feefinnegan.com

26/06/2020

# Contents

[Contents 1](#_Toc43488840)

[1. Introduction 2](#_Toc43488841)

[2. Project Scope and Objectives 3](#_Toc43488842)

[3. Student’s Learning Objectives 4](#_Toc43488843)

[4. Technical Specification of the Project 5](#_Toc43488844)

[5. Project Plan 6](#_Toc43488845)

[6. Conclusion 7](#_Toc43488846)

[7. References / Bibliography 8](#_Toc43488847)

# Introduction

*Provide some background information on the topic and introduce the importance of chosen topic. Highlight the problem/issue to be addressed.*

Over the last decade my career has centred on the management of software delivery projects in the financial crime prevention domain. In my opinion, it is clear that the industry is undergoing an evolution in software solutions, for regulatory compliance and fraud prevention, from ‘rules-engine’ technologies towards machine learning techniques.

A key motivator for me behind joining the DBS Higher Diploma in Science in Data Analytics was therefore to gain greater insight into this field of learning.

Another industry shift, which is obvious across multiple business domains, is the trend away from on-premises project deployments towards ‘cloud-based’ offering.

My project proposal is therefore a topic that I believe compliments these characteristics of the FinTech world by delivering a cloud hosted credit card fraud detection model. The model will be build and deployed via an Azure portal, but I will create a hosted Shiny R application to access the model in real time.

# Project Scope and Objectives

*Scope must state the range of functions/activities covered by the project. Objectives must be written with bullet points to list the steps required to achieve the functions mentioned as scope of the objects.*

The scope of this project can be summarised in the following points;

* An analysis of a Credit Card Fraud dataset (based on internal company data used for a, now discontinued, product line).
* The creation of a Credit Card Fraud Detection model, build through an Azure Machine Learning workspace.
* The deployment of a Rest endpoint within an Azure cluster/vm to assess the model.
* A Shiny R application, hosted within a separate Internet service to invoke the deployed model in real time with ‘new’ credit card data (this is data that is excluded from the training/testing of the model in Azure).
* I also intend to use the Shiny R App to perform a graphical analysis of the Credit Card dataset.

Project objectives can be summarised as;

* The successful training and testing of a one million row dataset in the Azure Machine Learning studio, with a resultant model that demonstrate a high degree of accuracy in credit card fraud detection.
* The deployment of an endpoint in Azure for which I can write R code routines to invoke a single API call to the model from an external application.
* This ‘external application’ will be an R Shiny application written by me for purposes of demonstrating the working deployment of the model.
* The Shiny R application will be hosted on-line and be the centre piece of the project demonstration. The application will also incorporate a visual analysis of the credit card dataset to exploit the richness R libraries to display data characteristics.
* Developing the Shiny R application is intended to show how a third party utility can access an Azure hosted model.

# Student’s Learning Objectives

*These objectives must show student’s self-motivated learning.*

I have determined the following personal objectives for this project;

* Credit card fraud detection through machine learning techniques is not an area with which I have worked previously. I wish to gain insight into how the process is performed.
* My company is investing heavily in Azure based services. I currently use some of these services for work in distributed PowerBI reports and VMs for software engineering development environments. The Machine Learning studio provided by Azure is a technology that greatly interests me and I want to use this project to dive deeper into some of it capabilities.
* In the first CA for the Advanced Data Analytics module (B8IT109) I experimented with using Shiny R applications and was impresses with the options to build visuals within browser based dashboards. I would like to continue to extend my knowledge of this technology through implementing a UI for this project.

# Technical Specification of the Project

Equipment and critical resources required to complete the project.

The project proposal has two major components to deliver;

* A deployed model for credit card fraud detection build in Azure Machine Learning Studio.
* A Shiny R application to act as both a UI for the model, and also to provide additional visual information on the structure of the Cred Card dataset used to build the model.

**Azure**

*The Azure Portal*

The Machine Learning Studio in azure is accessed through the Azure portal and, obviously, requires an Azure account in order to build and deploy the ML pipelines.

Options for access to Azure for me are either;

* Use existing DBS access, through which I have conducted my initial research.
* Use my company based account.

Azure services are not free but I have used a free educational account, via DBS, to date. If that becomes an issue I will revert to my company account, where I have a training allowance to complete the project.

*Azure Pipelines*

Model development will be through an ‘Azure Pipeline’. Work is likely to use the guidance of automated ML model generation utilities, followed by a visual model generation process – all of which is very analogous to project work I carried out in RapidMiner during the Data and Web Mining mdoeule (B8IT108).

I expect to need to embed Python or R code into the process, and may even switch to the Azure hosted Jypter Notebook approach if necessary. This is difficult to determine at the moment, but the Azure Machine Learning studio is very flexible in terms of SDK integration.

*Rest Endpoints*

The Azure Machine Learning Studio assists with code generation to invoke the deployed model through R, Python, or C# code. I have looked over a number of demo videos so I still expect the actual implementation of the code within my Shiny App to be challenging.

**Shiny R App**

The development on my Shiny R application will be a separate exercise, build in R Studio Cloud.

It will use open source third party libraries, such as Semantic Dashboard, to create an interesting visual user interface.

The application will be hosted externally, as opposed to running on a local machine. My intention is to research further into platforms like Shinyapps.io to host my project UI.

<Image> - Dev environments..

<image> - Deployed environments..

# Project Plan

A weekly schedule, indicating the proposed project’s ‘milestones’, which must be clear, concrete and achievable.

# Conclusion

Include concluding statements.

# References / Bibliography

References and bibliography must be written in Harvard referencing style.