Offer Acceptance Prediction Tool

Advanced Optimization Laboratory

Requirements Specification

Eric Le Fort

August 29, 2017

Contents

1 Introduction **2**
- 1.1 Project Overview . . . 2
- 1.2 Product Perspective . . . 2
  - 1.2.1 System Interfaces . . . 2
  - 1.2.2 Software Interfaces . . . 2
  - 1.2.3 Memory Constraints . . . 2
- 1.3 Description of Operation . . . 2
  - 1.3.1 Normal Operation . . . 2
  - 1.3.2 Error Handling . . . 2
- 1.4 Naming Conventions & Definitions . . . 3
  - 1.4.1 Definitions . . . 3
  - 1.4.2 Acronyms & Abbreviations . . . 3

2 Project Drivers **4**
- 2.1 The Purpose of the Project . . . 4
- 2.2 Clients, Customers, and Other Stakeholders . . . 4

3 Project Constraints **4**
- 3.1 Scope . . . 4
- 3.2 Relevant Facts & Assumptions . . . 4
  - 3.2.1 Facts . . . 4
  - 3.2.2 Assumptions . . . 4

4 Requirements **5**
- 4.1 Functional Requirements . . . 5
- 4.2 Non-Functional Requirements . . . 8
  - 4.2.1 Usability & Humanity Requirements . . . 8
  - 4.2.2 Performance Requirements . . . 9
  - 4.2.3 Operational & Environmental Requirements . . . 10
  - 4.2.4 Maintainability & Support Requirements . . . 11
  - 4.2.5 Security & Safety Requirements . . . 11
  - 4.2.6 Legal Requirements . . . 12

List of Tables

1 Revision History . . . 1
2 Definitions . . . 3
3 Acronyms and Abbreviations . . . 3

Date	Revision #	Comments	Authors
1-JUN-2017	0	- Initial document creation	Eric Le Fort
7-JUN-2017	1	- Document revision	Eric Le Fort Michael Liut
25-AUG-2017	2	- Revision midway through development	Eric Le Fort

Table 1: Revision History

1 Introduction

1.1 Project Overview

The core of this project involves developing a tool utilizing machine learning which can predict the number of enrolling students given an admission average or the admission average given a target number of enrolling students. Furthermore, developing this system will involve designing, constructing, and populating a database of relevant historic and current data.

1.2 Product Perspective

This section will outline various interfacing requirements as well as memory constraints of the system. Furthermore, it is important to note that this is intended as an improvement or extension of an existing project previously made by AdvOL.

1.2.1 System Interfaces

This system will require a user-level GUI which provides tools to perform predictions. This web interface will be designed to be compatible with both desktop and mobile devices.

1.2.2 Software Interfaces

This system must interface internally with its database of applicants. Furthermore, it will also need to interface with libraries which provide machine learning algorithms.

1.2.3 Memory Constraints

UNKNOWN - TO BE DETERMINED

1.3 Description of Operation

This section will describe both normal operation and how the system will handle unexpected events.

1.3.1 Normal Operation

Normal operation will consist of two main phases. The first phase will be training a new prediction model based on past data. This phase will occur as soon as the final data for the previous year is available. During this time, if an old model exists, the user can still access it but should be provided a visual warning that a new model is being trained. During the training phase, the system must first parse the applicant data provided by the university in order to store it in the database. Once it is in the database, the learning algorithm will use that data and available past data to train the system. The most recent year's data will remain for testing and as a control. As testing completes, the system will log the current model's accuracy for future reference.

The second phase will consist of using the trained model to predict results for the coming year. This will involve taking in either the target quantity of offers accepted or the minimum average desired and returning the expected value of the opposite parameter as determined by the trained model.

1.3.2 Error Handling

The main source of error in this project will arise while training the prediction model. If the prediction model is not able to converge to within its acceptable margin of error, this error should be logged in detail and a developer must work to modify the learning algorithm as necessary. The same type of process should be followed if the dimensionality reduction algorithm is unable to successfully reduce the dataset to within a reasonable feature count.

Another source of error will be due to the format of new data. This is an issue since this application has no control over the format of the data it is provided. In the case of an error of this nature, the error should be logged and a notification should be sent out to the developer indicating that they should check the error logs.

In the case of either type of error, a warning should be shown on the GUI to make the user aware that the new data has not been added or that the model might not be accurate.

1.4 Naming Conventions & Definitions

This section outlines the various definitions, acronyms, and abbreviations that will be used throughout this document in order to familiarize the reader prior to reading.

1.4.1 Definitions

Table 2 lists the definitions used in this document. The definitions given below are specific to this document and may not be identical to definitions of these terms in common use. The purpose of this section is to assist the user in understanding the requirements for the system.

Table 2: Definitions

Term	Meaning
Dimensionality Reduction Algorithm	This refers to an algorithm which selects a set of combined features which capture most of the variance of the data in an attempt to minimize the dimensionality of the feature space.
Availability Level	The percentage of time a system is available.
Machine Learning Algorithm	Any algorithm which uses past data in order to train a model to predict future data.
Prediction Model	A model built from a set of known information that can be used to predict a future event.
Train	The process of fine-tuning a prediction model using a machine learning algorithm based on past data.
Testing Data	Past data set aside to test the prediction model. This dataset must not overlap with the data used for training.
Training Data	Past data set aside to train the prediction model. This dataset must not overlap with the data used for testing.

1.4.2 Acronyms & Abbreviations

Table 3 lists the acronyms and abbreviations used in this document.

Table 3: Acronyms and Abbreviations

Acronym/Abbreviation	Meaning
AdvOL	Advanced Optimization Laboratory
GUI	Graphical User Interface
EOAPT	Engineering Offer Acceptance Prediction Tool
IRA	Office of Institutional Research and Analysis
VPN	Virtual Private Network
KPCA	Kernel Principal Component Analysis
MIS	Module Interface Specification
MID	Module Internal Design

2 Project Drivers

2.1 The Purpose of the Project

The purpose of this project is to predict the amount of individuals that will accept an offer of enrolment from McMaster University in the faculty of engineering given an admission average or the admission average given a number of enrolling students. This tool will greatly assist the Dean's office in deciding what the minimum required grade average for candidates should be.

2.2 Clients, Customers, and Other Stakeholders

The client for this project is the Dean of Engineering, Dr. Ishwar Puri, and Assistant Dean, Maria White, for the faculty of engineering at McMaster University.

Future customers will include the admission departments for the rest of McMaster University.

Stakeholders of this project include the faculty of engineering and the department of admissions at McMaster University as well as the IRA, AdvOL, and applicants to the programs hosted by McMaster University.

3 Project Constraints

These are the constraints that are necessary for the operation and development of this program:

- All identifying data must remain confidential and protected.
- The application cycle must be adhered to.

3.1 Scope

The scope of this project will currently only take into account students entering either the co-op or non-co-op engineering programs. Further prediction models would have to be introduced in order to support other programs as well. The only data used will be that which is provided to AdvOL by the university (i.e. the team will not work on gathering its own data).

3.2 Relevant Facts & Assumptions

This section will outline various facts and assumptions that will be leveraged in the design of this system.

3.2.1 Facts

- The format of data of previous years.
- There will be unknowable underlying data that is important to an applicant's final choice.
- This model will not account for factors such as political, environmental, or personal influences.

3.2.2 Assumptions

- Future data formats will not deviate that far from past formats.
- Rates of grade inflation will be predictable.
- Recent applicants are more similar to current applicants than less recent ones.

4 Requirements

The following section outlines the various requirements of this system. For each requirement, a description of the requirement, the rationale behind the requirement, and a fit criterion for when the requirement will be satisfied are provided. The values given for customer satisfaction and dissatisfaction are from 0-5. The values for requirement priority are either high, medium or low. All high priority requirements are necessary for the system to be functional. Failing to meet too many medium priority requirements may result in considering the system unsatisfactory. Low priority requirements would be nice to have but are not mission critical.

4.1 Functional Requirements

The following is a collection of the functional requirements of this system:

Requirement #: 1 **Requirement Type**: F

Description: The system must be able to add applicant data to the database.
Fit Criterion: New data is added in large batches in a way that is not time consuming and there is no data missing or inputted incorrectly.

Originator: Eric Le Fort

Customer Satisfaction: 1 **Customer Dissatisfaction**: 5
Priority: HIGH **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 2 **Requirement Type**: F

Description: The system should be able to perform the dimensionality reduction algorithm on the full dataset.
Fit Criterion: The features automatically selected maximize the variance in the data using twenty or fewer features.

Originator: Eric Le Fort

Customer Satisfaction: 1 **Customer Dissatisfaction**: 1
Priority: MEDIUM **Conflicts**: P1, P3

History: Created 1-JUN-2017

Requirement #: 3 **Requirement Type**: F

Description: The system must be able to train and re-train the prediction model as necessary.
Fit Criterion: The system can train a prediction model given at least one year of data.

Originator: Eric Le Fort

Customer Satisfaction: 2 **Customer Dissatisfaction**: 5
Priority: HIGH **Conflicts**: OE2

History: Created 1-JUN-2017

Requirement #: 4 **Requirement Type**: F

Description: The system must be able to predict the quantity of accepted offers based on a given minimum grade average.
Fit Criterion: The system can predict the quantity of accepted offers based on a given minimum grade average.

Originator: Eric Le Fort

Customer Satisfaction: 5 **Customer Dissatisfaction**: 5
Priority: HIGH **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 5 **Requirement Type**: F

Description: The system must be able to predict the minimum grade average required based on a given target quantity of accepted offers.
Fit Criterion: The system can predict the minimum grade average required based on a given target quantity of accepted offers.

Originator: Eric Le Fort

Customer Satisfaction: 5 **Customer Dissatisfaction**: 5
Priority: HIGH **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 6 **Requirement Type**: F

Description: The system must cache previous prediction results to minimize the computation time of subsequent requests.
Fit Criterion: After any prediction is performed, the system caches that result. Also, if a cached result is available, the system uses that result.

Originator: Eric Le Fort

Customer Satisfaction: 1 **Customer Dissatisfaction**: 3
Priority: LOW **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 7 **Requirement Type**: F

Description: The system must clear the cache of previous prediction results when they are no longer valid (i.e. the prediction model or data is altered). This requirement is contingent on requirement 7 being implemented.
Fit Criterion: When a new user model is trained, the cache has been cleared.

Originator: Eric Le Fort

Customer Satisfaction: 1 **Customer Dissatisfaction**: 5
Priority: HIGH **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 8 **Requirement Type**: F

Description: The system must perform previously unperformed predictions for various input parameters while the system is idle. This will maximize the responsiveness of the system to user requests.
Fit Criterion: If idle for an hour, the system begins performing previously unperformed predictions.

Originator: Eric Le Fort

Customer Satisfaction: 3 **Customer Dissatisfaction**: 2
Priority: LOW **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 9 **Requirement Type**: F

Description: The system must allow the user to interactively choose their input parameter type (minimum grade average or target acceptance quantity) and corresponding value.
Fit Criterion: The user interface allows dynamic selection of input parameter type.

Originator: Eric Le Fort

Customer Satisfaction: 5 **Customer Dissatisfaction**: 5
Priority: HIGH **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 10 **Requirement Type**: F

Description: The system must visually display results of the prediction to the user.
Fit Criterion: The user interface displays results of the prediction.

Originator: Eric Le Fort

Customer Satisfaction: 5 **Customer Dissatisfaction**: 4
Priority: HIGH **Conflicts**: None

History: Created 1-JUN-2017

4.2 Non-Functional Requirements

The following is a collection of the non-functional requirements of this system:

4.2.1 Usability & Humanity Requirements

Requirement #: 1 **Requirement Type**: UH

Description: The GUI must be simple to use.
Fit Criterion: This requirement will be considered satisfied if a user without training can perform a usage scenario within 1 minute of being introduced to it.

Originator: Eric Le Fort

Customer Satisfaction: 5 **Customer Dissatisfaction**: 3
Priority: LOW **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 2 **Requirement Type**: UH

Description: The information provided to the user must be straightforward and require minimal interpretation.
Fit Criterion: This requirement will be considered satisfied if a user not familiar with the system or the specific context is able to understand all information the interface provides.

Originator: Eric Le Fort

Customer Satisfaction: 5 **Customer Dissatisfaction**: 5
Priority: LOW **Conflicts**: None

History: Created 1-JUN-2017

4.2.2 Performance Requirements

Requirement #: 1 **Requirement Type**: P

Description: The prediction model must be able to be trained quickly to handle new data.
Fit Criterion: This requirement will be considered satisfied if the prediction model can train within a day using modest computing resources.

Originator: Eric Le Fort

Customer Satisfaction: 2 **Customer Dissatisfaction**: 3
Priority: MEDIUM **Conflicts**: F3, P3

History: Created 1-JUN-2017

Requirement #: 2 **Requirement Type**: P

Description: The system must be able to provide results quickly in order to allow the user to experiment with different inputs.
Fit Criterion: This requirement will be considered satisfied if the system performs its predictions, using a trained prediction model, within 10 seconds of a user's request.

Originator: Eric Le Fort

Customer Satisfaction: 4 **Customer Dissatisfaction**: 5
Priority: MEDIUM **Conflicts**: MS1, MS2

History: Created 1-JUN-2017

Requirement #: 3 **Requirement Type**: P

Description: The system must provide accurate predictions.
Fit Criterion: This requirement will be considered satisfied if the maximum error between predicted and actual acceptance is no more than 5%.

Originator: Eric Le Fort

Customer Satisfaction: 5 **Customer Dissatisfaction**: 5
Priority: MEDIUM **Conflicts**: F3, P1

History: Created 1-JUN-2017

4.2.3 Operational & Environmental Requirements

Requirement #: 1 **Requirement Type**: OE

Description: The system must minimize client-side demand.
Fit Criterion: This requirement will be considered satisfied if a low-tier smartphone (i.e. a phone released about 5 years ago) is able to handle any client-side computational demand of this system without a noticeable performance penalty (a half second or longer delay) as compared to a more high-end device.

Originator: Eric Le Fort

Customer Satisfaction: 3 **Customer Dissatisfaction**: 5
Priority: LOW **Conflicts**: None

History: Created 1-JUN-2017

Requirement #: 2 **Requirement Type**: OE

Description: The system must be available for the user as often as possible.
Fit Criterion: This requirement will be considered satisfied if the system is able to achieve an availability level of 99.5%.

Originator: Eric Le Fort

Customer Satisfaction: 3 **Customer Dissatisfaction**: 5
Priority: LOW **Conflicts**: F4

History: Created 1-JUN-2017

4.2.4 Maintainability & Support Requirements

Requirement #: 1 **Requirement Type**: MS

Description: New data will be added every year as new applicants apply. The system should be able to integrate this new data into the database. This is important since the product will likely outlive the development team.
Fit Criterion: This requirement will be considered satisfied if the system can handle the current range of input data.

Originator: Eric Le Fort

Customer Satisfaction: 1 **Customer Dissatisfaction**: 1
Priority: MEDIUM **Conflicts**: None

History: Created 1-JUN-2017

4.2.5 Security & Safety Requirements

Requirement #: 1 **Requirement Type**: S

Description: The system should be accessible by any approved user with VPN access to the internet.
Fit Criterion: If the system is online, any user with the necessary permissions is granted access.

Originator: Eric Le Fort

Customer Satisfaction: 1 **Customer Dissatisfaction**: 5
Priority: HIGH **Conflicts**: P2

History: Created 1-JUN-2017

Requirement #: 2 **Requirement Type**: S

Description: The system must not be accessible by any unapproved user.
Fit Criterion: If an unauthorized user attempts to access the system, their request is denied.

Originator: Eric Le Fort

Customer Satisfaction: 1 **Customer Dissatisfaction**: 1
Priority: HIGH **Conflicts**: P2

History: Created 1-JUN-2017

4.2.6 Legal Requirements

Requirement #: 1 **Requirement Type**: L

Description: The system must adhere to all applicable data governance laws it falls under.
Fit Criterion: The system adheres to all applicable data governance laws it falls under.

Originator: Eric Le Fort

Customer Satisfaction: 1 **Customer Dissatisfaction**: 1
Priority: HIGH **Conflicts**: None

History: Created 1-JUN-2017