**Machine Learning Using Deep Learning and Natural Language Processing**

**Peraton**

**www.peraton.com**

**Requirements**

**Date:**

09/24/2021

**Team Members:**

Rolando Gonzalez

Elshaday Alemayehu

Yusuf Siddiqui

Arial Carvalho

Eduardo Heredia

# Abstract

This document details the agreement between the sponsor and the development team. The purpose of this document is to provide the outcome from requirements elicitation between the development team and the sponsor. It focuses on the functional requirements, non-functional requirements, and the use case model of the application. This document is not final and is subject to change as needed by the sponsor.

The functional requirements define the intended system behavior and provide a high-level overview of the system’s capabilities. The non-functional requirements define the ‘how’ of the system; the system shall meet certain targets to satisfy the security, usability, maintainability, scalability, and availability aspects of the system. The use case of the system shows how the user will interact with the system and how the system will satisfy the basic needs of the sponsor. Each use case is documented in detail under the textual description section and the rationale is provided for the use case model.

# Table of Contents

[Abstract 2](#_Toc83392315)

[Table of Contents 3](#_Toc83392316)

[List of Figures 4](#_Toc83392317)

[Introduction 5](#_Toc83392318)

[System Use Case Model 6](#_Toc83392319)

[Graphic Use Case Model 6](#_Toc83392320)

[Textual Description for Each Use Case 7](#_Toc83392321)

[Non-Functional Requirements 11](#_Toc83392322)

[Evidence the Requirements Have Been Placed Under Configuration Management 12](#_Toc83392323)

# List of Figures

Figure 1: System Use Case 6

# Introduction

This requirements document presents a series of requirements that are crucial towards defining the needs of the project. Some details include the functional requirements, non-functional requirements, use case model, descriptions for each use case, and the rationale behind the use case model. The tracking of this information is crucial for the successful deliverable of the project ensuring all three main pillars are met: budget, schedule, and scope.

The purpose of functional requirements is to ensure the project has a clearly defined scope and that the needs of the sponsor align with the plans of the development team.

The project sponsor is responsible for defining the scope of the project, managing all the modifications related to scope, and approving the deliverables for scope changes.

The product to be developed is a program with capability of determining the quality and maturity of documents based on predetermined hypotheses. The program shall read then process the documents using NLP technologies, and accomplish its goal using ML/AI models. The underlying motivation for this product is to improve decision-making process as determining the level of risk a contract may impose.

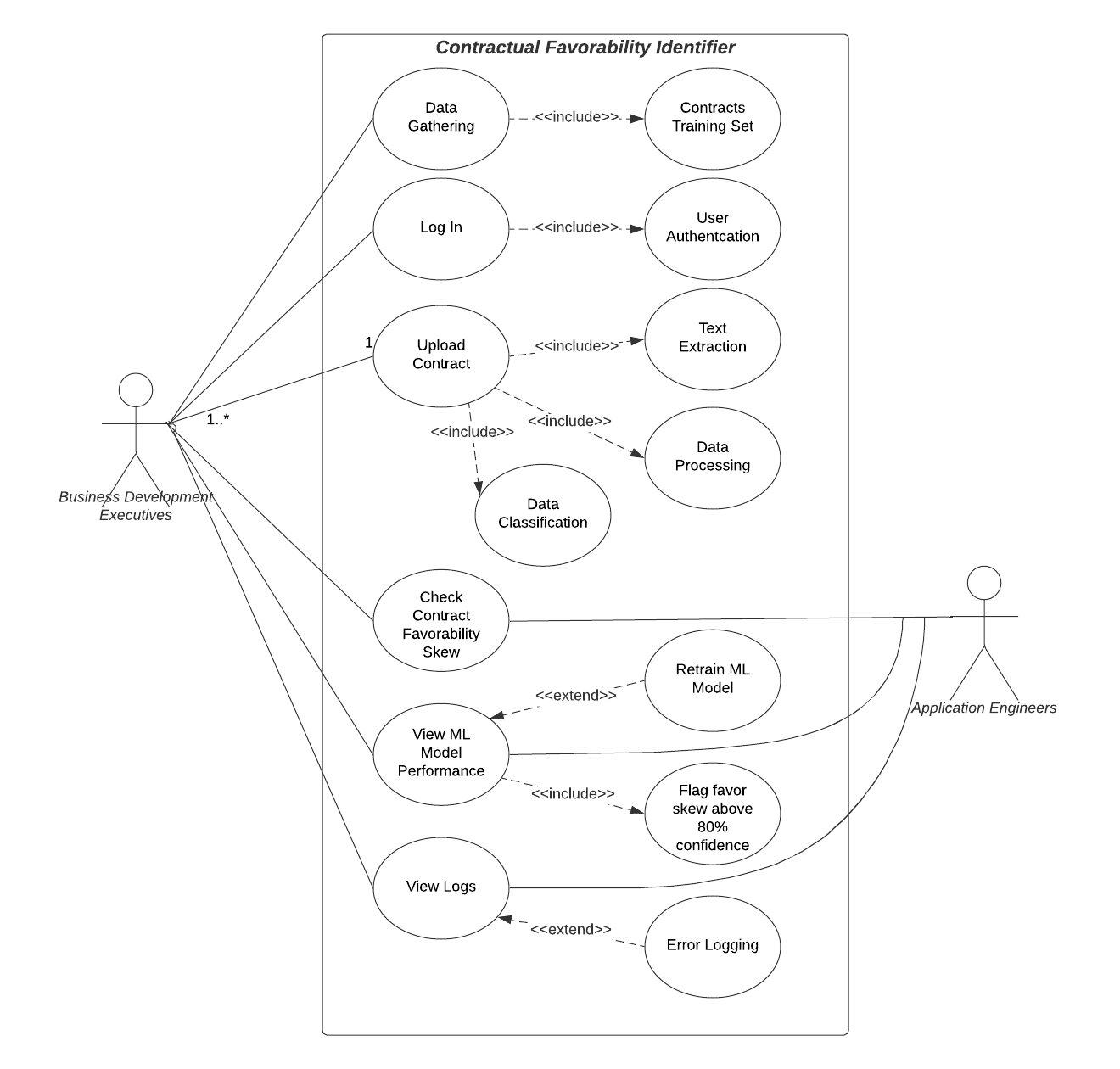
The structure of the requirements document is as follows:

1. The use case model which shows the system users and how they interact with the system.
2. Textual description of each use case providing use case name, participating actors, entry conditions, flow of events, exit conditions, exceptions, special requirements
3. Rationale for use case model which provide a summary of the logic behind the use cases and interactions
4. Non-functional requirements which indicate system goals in terms of performance, security, maintainability, usability, and availability
5. Configuration management which allows for the document to be modified if requirements change throughout the project

# System Use Case Model

## Graphic Use Case Model

Figure : System Use Case



## Textual Description for Each Use Case

R01: Account Creation

Participating Actors: Business Development Executive

Entry condition: User has/knows an email address, password, and name

Flow of events:

1. User clicks on create account.
2. System displays create an account page.
3. User enters their account information.
4. System validates user information.
5. System persistently stores user data to the database.
6. System displays a message that an account has been created.

Exit conditions: User creates an account to access the system

Exceptions:

1. User clicks on create account.
2. System displays sign up page.
3. Users enter their information.
4. System identifies entered data that is already stored in the database.
5. System displays a message that an account with the data provided is in use.

Special Requirements: Once all required information is entered by the user, the system shall validate and displays the result within 5 seconds.

R02: **Log In**

Participating actors: Business Development Executive

Entry condition: User has an existing account created within the system.

Flow of events:

1. User clicks on log in
2. System displays login page
3. User enters their login information
4. System validates user information
5. System displays main menu from the dashboard

Exit conditions: User is granted access to the system main menu.

Exceptions:

1. User clicks on log in
2. System displays login page
3. User enters their information
4. System identifies entered data is not stored in database or incorrect
5. System displays a message that user has entered an incorrect login credential

Special Requirements: Once all required information is entered by the user, the system shall validate and displays the result within 5 seconds

E01: User sets system to their preferred language

Participating Actors: Business Development Executive

Entry condition: User has an option to choose from multiple languages

The flow of events:

1. User logs in to account
2. System asks the preferred desired language
3. System asks for document input
4. User enters information
5. System displays result in user preferred language

Exit conditions: System sets the chosen language as default for the user

Exceptions:

1. User logs in to their account
2. The system sets the default language to the user chosen language

Special Requirements: After the user chooses the language, the results should be converted to the preferred language within 5 seconds.

E02: **Upload Contract** - User shall be able to upload desired contract in pdf file format

Participating Actors: Business Development Executive

Entry condition: Users will be able to input the contract in the pdf file format

The flow of events:

1. User logs in to account
2. System asks to input desired document
3. The user enters pdf format documents
4. System checks if the document is in the correct format
5. The system processes the contract
6. System displays results

Exit conditions: The user enters a different document format

Exceptions:

1. User logs in to their account
2. System asks to input desired document
3. The user enters documents
4. System checks if the document is in the correct format
5. The system notifies the user to enter the correct format

Special Requirements: The system should be able to verify the document type in less than 3 seconds

E03: **Text Extraction** - The system must be able to extract text from user input documents

Participating Actors: Business Development Executive, System

Entry condition: In order to categorize the given contract into relevant dataset, the system will be able to complete text extraction from a given document

The flow of events:

1. Ask the user for input
2. The user enters pdf format documents
3. The system extracts text from the pdf file
4. The system cleans the documents
5. The system prepares text for a further use

Exit conditions: Text is extracted and organized properly

Exceptions: User aborts the process

Special Requirements: The system should be able to extract text from the document within minutes

E04: Upload contract - Data processing

Participating Actors:  Business Development Executive, Application Engineers

Entry condition: The system will be able to clean and convert raw data into a usable format

The flow of events:

1. System cleans text and split it into word
2. System processes the data
3. System categorizes the raw data into datasets
4. System stores the contract data into a dataset for further use

Exit Conditions: Raw data is converted to an applicable Machine Learning format

Exceptions: System not able to find keywords from the given document

Special Requirements: The system should be able to process the document within hours

Y01: **Check Contract Favorability Skew** – User shall be able to view flagged keywords and phrases

Participating Actors: Business Development Executive, Application Engineers

Entry condition: The ML model shall determine which keywords and phrases and provide insight on the favorability of flagged information

The flow of events:

1. User tunes acceptable parameters for ML model
2. User inputs document
3. System extracts text from the document
4. System processes the data
5. System classifies the data
6. System organizes data into charts and tables and displays information on the dashboard

Exit conditions: The user leaves parameters blank for ML model

Exceptions: User terminates the process early

Special Requirements: The system should be able to process the document within hours

Y02: **View ML model performance** – User shall be able to view overall performance of the ML model

Participating Actors: Business Development Executive, Application Engineers

Entry condition: The ML model shall provide information regarding words that did not meet the threshold parameters specified by the user.

The flow of events:

1. User enters bounds for confidence level and modifies settings
2. User inputs document
3. System provides information within specified bounds

Exit conditions: The user leaves parameters blank

Exceptions: User enters invalid bound range

Special Requirements: The system should be able to retrieve information within specified bounds

Y03: **View Logs** – System shall provide logs regarding information about the application such as successful logins, failed logins, errors, timestamps, warnings, and critical failures.

Participating Actors: Business Development Executive, Application Engineers

Entry condition: User is a valid elevated user within the system and has successfully logged in

The flow of events:

1. System prompts user for login information, title, employee id
2. User enters information correctly
3. User requests to view system logs
4. System provides the user with logs within an entered range (date, time, etc.)

Exit conditions: User does not have permission to access system logs

Special requirements: The system should be able to store and maintain logs for the user pertaining to the past 1 year.

Rationale for Use Case Model:

The system should be able to parse and extract data to fit a ML model and should allow for retraining of data to be able to adapt to poor accuracy. The system shall also be able to provide detailed analysis of the requested information in the form of graphs and tables. The system shall be able to provide system logs to allow for monitoring and debugging.

# Non-Functional Requirements

Security

* Passwords must be hashed and stored securely in the database
  + The application will not be available to all users, valid users must authenticate and verify identity with their title/position, employee id (if applicable), and password
* Sensitive information shall be encrypted for in-flight data
  + Information contained within the contracts that the program works with may be sensitive and will require some form of encryption when the application is run over a network
* System will automatically sign out within 15 minutes if no work is in progress and no response from the user is detected
  + Login session will be terminated to prevent sensitive data from being leaked when users are no longer using the application and no work is running
* Passwords must contain at least 8 characters including at least one special character, upper case letter, and number
  + Valid users will need to create a strong password as the program will deal with sensitive information
* Password delay after 10 consecutive failed login attempts
  + Too many failed login attempts will require the user to verify their identity or wait a certain amount of time to prevent brute-force attempts
* The user account database must use column-level transparent data encryption or better
  + A database that holds login information is sensitive and needs to be adequately encrypted in order to be properly protected
  + Require every column in the database a password for authorizing access to column data

Maintainability

* The system shall allow elevated users to provide a new training set of contractual documents
  + When there is high deviation from the expected outcome and the ML model’s accuracy is performing poorly, the model may need to be retrained with a new set of documents

Usability

* The system must work on OSX, Windows, and Unix-based systems
  + Even though development is done in a Windows environment, the dashboard should still be usable in other operating system environments
* The system may support multiple languages
  + A possible add-on would be to support contracts that may be in other languages
* The system must provide sorting functionality
  + Sorting should be provided for highest confidence levels, favorable/unfavorable keywords, etc.

Availability

* The system must retrieve contract documents one at a time
  + A user will be able to input one pdf document at a time instead of allowing users to input multiple files at a time.
  + Limit user input to only one at a time
* The system should not take longer than 3 seconds to log in the user
  + The system should validate user log-in by cross checking email and password and this should be complete in less than 3 seconds
* The system must be able to process contract analysis within hours

# Evidence the Requirements Have Been Placed Under Configuration Management

