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

**Peraton**

**www.peraton.com**

Project Management Plan

**Date:**

09/10/2021

**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 a high-level overview of the development process needs and features. It focuses on the capabilities needed by the sponsor and why those needs exist. This document outlines the project plan, scope of the project, and deliverables. This document is not final and is subject to change as needed by the sponsor.

The development team is highlighted under the project organization and demonstrates each member’s responsibility to ensure timely completion of deliverables. The lifecycle model demonstrates the approach in which phases of the project will be completed. The risk analysis highlights potential issues that may be faced during development and mitigation strategies. The hardware and software resource requirements list the basic needs of the development team to develop the product. The schedule shows the estimated time to complete deliverables and is subject to change based on the needs of the sponsor. The monitoring, reporting, and controlling mechanisms outline the reports that should be produced. Professional standards section lists what is expected of each team member during the development of this project.

Table of Contents

[Abstract 2](#_Toc82184419)

[List of Tables 4](#_Toc82184420)

[Introduction 5](#_Toc82184421)

[Project Organization 6](#_Toc82184422)

[Lifecycle Model Used 7](#_Toc82184423)

[Development Lifecycle Model 7](#_Toc82184424)

[Machine Learning Lifecycle Model 7](#_Toc82184425)

[Risk Analysis 8](#_Toc82184426)

[Hardware and Software Resource Requirements 10](#_Toc82184427)

[Deliverables, Schedule 11](#_Toc82184428)

[Monitoring, Reporting, and Controlling Mechanisms 12](#_Toc82184429)

[Professional Standards 13](#_Toc82184430)

[Acronym Table 14](#_Toc82184431)

[References 15](#_Toc82184432)

[Appendix A. 15](#_Toc82184433)

# List of Tables

[Table 1: Team Responsibilities 6](#_Toc82183618)

[Table 2: Project Issues and Risks 8](#_Toc82183619)

[Table 3: Deliverables and Schedule 11](#_Toc82183620)

[Table 4: Acronym Table 14](#_Toc82183621)

# Introduction

This project plan document presents a series of scheduled activities and important details for the development of the project. Some details include the project governance, potential risks, software and hardware specifications, project deliverables, and professional standards. 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 the Scope Management Plan is to ensure the project pertains to the requirements and objectives of the work required to complete the project successfully. The scope document also states the methodology we are applying to manage the scope of the project, determining the roles and responsibilities for the project, and how we are going to validate and control the scope of the project.

The following points are the different aspects the scope management plan has:

* Collection of requirements of Peraton for the AI/ML program
* Scope definition of AI/ML program
* Scope Validation
* Scope Control

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 and process documents using NLP technologies, and accomplish its goal by 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 project plan will follow the process of the ML lifecycle and SDLC as described in detail further below.

# Project Organization

Table 1: Team Responsibilities

|  |  |  |
| --- | --- | --- |
| **Team Member** | **Role** | **Responsibilities** |
| Rolando Gonzalez | Project Lead | 1. Be the representative and intermediary between the team and the sponsors 2. Ensure team is meeting weekly 3. Keep track of activities performed and deliverables 4. Communicate project stakeholders and ensure deliverables are meeting deadlines |
| Ariel Carvalho | ML/AI Developer | 1. Design ML models 2. Research and implement ML/AI algorithms 3. Select and test data set 4. Collect and verify ML model results are accurate |
| Elshaday Alemayehu | NLP Specialist | 1. Design NLP systems 2. Use effective text representation to transform natural language into useful features 3. Implement accurate algorithms and tools for NLP tasks 4. Test and validate NLP requirements |
| Yusuf Siddiqui | ML/AI Developer | 1. Design ML models 2. Research and implement ML/AI algorithms 3. Select and test data set 4. Collect and verify ML model results are accurate |
| Eduardo Heredia | Front-end/UX Developer | 1. Create prototypes and wireframes 2. Implement and test user-friendly design 3. Develop responsive user interfaces 4. Validate input and output from users |

The team has been given their role based on skills and previous experience. The motivation for the distribution is to certify all main areas of the ML/AI program are being assigned for at least one team member.

# Lifecycle Model Used

## Development Lifecycle Model

Our project would be using a Waterfall life cycle model that involves the following major steps:

* Requirement

The team will be able to analyze the requirements to fully understand the problem and document them properly.

* Design

In this phase, the team will take the requirement specification and the system design will be prepared. This system design helps us in specifying hardware and system requirements.

* Implementation

In this phase, we will begin technical implementation based on the system design.

* Testing

All the units developed in the implementation step will be integrated into a system after testing each unit.

* Deployment

Once the functional and non-functional testing is done,  the product will be deployed in the Peraton production environment.

## Machine Learning Lifecycle Model

Our project would be using Machine learning life cycle that involves the following major steps:

* Gathering Data
  + In this first step of the machine learning life cycle, the team will identify different data sources that can be collected from different sources such as files, documentation, and company contracts. By integrating the data obtained from different sources, the team would be able to get a coherent set of data.
* Data Preparation
  + After getting datasets, we put our data in the right place and prepare it to use in our machine learning training. This can be done using data exploration to understand the format and characteristics of the data for a better understanding of the data that can lead to an effective outcome.
* Data Processing
  + In this ML lifecycle step, our team will be cleaning and converting raw data into a usable format using natural language processing techniques.
* Analyze Data
  + The team will build machine learning models to analyze the data using different machine learning techniques and evaluate the model.
* Train and Test the Model
  + The team will use datasets to train the model using different machine learning algorithms approaches. This step will help us to understand the various patterns and features within our data. After training, we will determine the percentage accuracy of the model by testing against the trained data.
* Deployment
  + Finally, we will deploy the model in the real-world system.

# Risk Analysis

Table 2: Project Issues and Risks

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| High Risk |  |  |  |  |  |
| Med Risk | **Project - Issues and Risks** | | | | |
| Low Risk | **Item #** | **Description** | **Resolution Plan / Mitigation Plan** | **Contingency Plan** | **Risk Owner** |
|  | R-1 | Unplanned work that must be accommodated | Attend project scheduling meetings.  Document all assumptions made in planning and communicate to the project manager before project kick off. | Escalate to the team leader with plan of scheduling, including impact on time, cost, and quality | Team Leader |
|  | R-2 | Project may experience delays due to code development components being built from scratch. This may consume time and effort, compared to reusing codes that already exist. | appoint more resources to the developing team.  Code reuse, developers are able to use stable and trusted code libraries, best for building blocks  Open Source | using Open Source for building block in implementation of code | Developers |
|  | R-3 | Project may experience delay due to lack of expertise in using test frameworks. This might be a challenge for software testers to complete their task in time. | Hand the task to team members with more testing skills to learn more about the required test framework to help with the task | Give team members open source learning resources to develop such skill in short time | Developers |
|  | R-4 | Project may experience delays due to incompatibility issues that might arise on production environments which the development teams haven’t encountered on non-production environments. | Involve experienced devOps engineers to help prepare for these kinds of problems  Replicate the issues on non-production environments to help mitigate the issues | Communicate with production environment support team for possible solution to the incompatibility problems | Developers |
|  | R-5 | Project schedule is not clearly defined or understood | Share the schedule and go through upcoming tasks at  each weekly project progress meeting. | Revisit the schedule with team leader and relaunch the new project schedule | All Team members |
|  | R-6 | Lack of communication, causing lack of clarity and confusion. | Write a communication plan that includes: the frequency, goal, and audience of each communication.  Identify stakeholders early and make sure they are considered in the communication plan. | Correct misunderstandings immediately.  Clarify areas that are not clear swiftly using assistance from Project Sponsor if needed. | Team lead |
|  | R-7 | Project might be delayed due to late arrival of necessary documentation from stakeholders | Set up a meeting with stakeholders to discuss details about necessary resources and documentation | Start on other tasks that are not dependent on the specific documentations | Team lead |
|  | R-8 | Project design and deliverable definition is incomplete | Define the scope in detail via meetings and include inputs from subject matter expertise | Document assumptions made and associated risks | Project Sponsor |

# Hardware and Software Resource Requirements

Hardware: Windows laptop environment with Wi-Fi connectivity

Software: Python, Anaconda, Jupyter Notebook

* All team members will be operating in a Windows environment to avoid conflicts
* We will be using Python as it is suitable for collaborative implementation within our team members. Since Python is a general-purpose language, it can do a set of complex machine learning tasks and enable us to build prototypes quickly that allow us to test our product for machine learning purposes. Python also has a lot of ML and NLP  frameworks and libraries that can be suitable for our project.
* We will also be using Jupyter notebooks for all sorts of data tasks including data cleaning and transformation, data visualization, machine learning, deep learning, and NLP.

# Deliverables, Schedule

Table 3: Deliverables and Schedule

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Activity** | **Dependency** | **Estimated Completion** | **Team Members** |
| A01 | Project Management Plan | N/A | 09/10/2021 | All |
| A02 | Requirements Gathering | A01 | 09/24/2021 | All |
| A03 | Architecture Documentation | A02 | 10/08/2021 | All |
| A04 | Design Documentation | A03 | 10/22/2021 | All |
| A05 | Data Gathering | A02 | 10/29/2021 | Rolando Gonzalez |
| A06 | Developing Algorithms for Text Extraction | A05 | 11/03/2021 | Elshaday Alemayehu |
| A07 | Data Processing(clean/prepare data) with NLP | A06 | 11/06/2021 | Elshaday Alemayehu |
| A08 | Classification | A07 | 11/08/2021 | Eduardo Heredia |
| A09 | Training ML Model | A08 | 11/10/2021 | Ariel Carvalho  Yusuf Siddiqui |
| A10 | Testing Plan | A09 | 11/12/2021 | All |
| A11 | Evaluating ML Model | A10 | 11/22/2021 | Ariel Carvalho  Yusuf Siddiqui  Eduardo Heredia |
| A12 | Final Report | A11 | 12/03/2021 | All |

The project will follow the above schedule to fit both of the chosen lifecycle models used in the project. This includes the deliverables required and the high-level milestones needed to successfully complete the project.

# Monitoring, Reporting, and Controlling Mechanisms

* To set up our project monitoring and control process, we first are going to establish the project baselines such as scope, schedule, and budget. Using this information team members shall create a work breakdown structure (WBS) to break down the project into small units of work and sub-tasks. WBS will help us detect issues at an early stage which will be important to keep the project under control.
* The team will be using a requirement traceability matrix to ensure all requirements are defined, accepted requirements are broken down into development tasks and changes at any time during the development lifecycle are traced.
* The team will be using RAID to identify project risk, assumptions, issues, and dependencies. RAID will help us proactively manage the project.
* To monitor and control the project, we will be using project schedules to help us assist with project tracking, reporting progress, identifying task relationships, monitoring progress, and identifying issues early.
* Governance Management Plan

# Professional Standards

* Each team member will accept full responsibility for their own work.
* Each team member will place great effort on the product and their related work to meet high quality.
* Each team member will support other team members.
* Each team member shall use their professional judgment to make informed decisions on their work.
* Each team member shall produce their own original work.
* The team shall display the best interest to their employer (Peraton).
* Team members missing meetings, deliverables, etc., will need to provide valid reasons to be excused.
  + A valid reason as an example:
    - Illness, family member’s death, travels, academic reasons, etc.
  + An invalid reason as an example:
    - “I forgot,” “device is drained,” “I am busy with something else,” etc.
* Team members who show misbehavior will be evaluated by the team and instructor. (This refers to appendix A)
  + 1st occurrence, the situation will be evaluated and the team shall resolve the underlying problem.
  + 2nd occurrence, in addition to the before, the instructor will be notified about the problem, and the team will meet with the instructor to resolve the problem.
  + 3rd occurrence, the professor will be notified, and the same conduct will occur. In addition, the team will be allowed to evaluate and decide whether the team member should be removed.
  + Each occurrence will be documented.
  + Examples of misbehavior may include but not limited to:
    - Missing deliverables, missing team meetings, poor quality of work, disrespectful attitude, harassment, unprepared for meetings, neglectful acts, etc.

All team members are subject to the above standards to ensure the project progresses smoothly and failure to comply with these rules will result in documentation of the incident and notifying the instructor if appropriate.

# Evidence the Document Has Been Placed Under Configuration Management

Graphical user interface, text, application, email

Description automatically generated

# Acronym Table

Table 4: Acronym Table

|  |  |
| --- | --- |
| **Acronym** | **Meaning** |
| ML | Machine Learning |
| AI | Artificial Intelligence |
| NLP | Natural Language Processing |
| UX | User Experience |
| RAID | Risk, Assumptions, Issues, Dependency |
| SDLC | Software Development Life Cycle |

# References

# Appendix A.

* The following provides a professional standards guideline for the teams. This guideline may be tailored. The professional standards must be agreed upon by each member in the team.
* Guidelines:
  + On the first occurrence of unacceptable behavior, determine the circumstances involved, resolve the problem, and document the event in the meeting minutes.
  + On a second occurrence, notify the instructor of the problem. A meeting will be set up to evaluate the situation and resolve the problem.
  + On a third occurrence, again notify the instructor of the problem. A meeting will be set up to evaluate the situation and resolve the problem. At this point, the team will have the \*option\* of removing the team member. If removed, then the team member receives a prorated grade based on the number of weeks they have participated in the group.
  + Examples of unacceptable behavior may include not delivering on time, delivering poor quality work, missing team meetings, being unprepared for team meetings, disrespectful or rude behavior, etc.
  + Reasons such as "too busy" or "I forgot", or "my dog ate my design model" are unacceptable.
  + Valid reasons that must be considered include those listed for obtaining an incomplete standing in a course (illness, death in the family, travel for business or academic reasons, etc.)