

NIP2 – NIP2 TASK 3: MACHINE LEARNING PROJECT PROPOSAL

INTRODUCTION TO ARTIFICIAL INTELLIGENCE – C951

PRFA – NIP2

TASK OVERVIEW

SUBMISSIONS

EVALUATION REPORT

COMPETENCIES

4036.1.11: Machine Learning

The graduate creates models with machine learning algorithms in order to extract actionable insights from data.

INTRODUCTION

Machine learning is the “science and art of programming computers so they can learn from data” (Géron, 2019, p. 1). Machine learning is a collection of approaches that automates analytical model building and supports the discovery of insights from data. This assumes that algorithms can learn and improve from data, including examples, direct experience, or instruction; identify relationships; and help make decisions with minimal interventions from the human user. It is the pivotal engine behind artificial intelligence efforts supporting supervised, unsupervised, and reinforced learning algorithms enabling the analysis of massive amounts of data.

While early forms of machine learning have existed for decades, in recent years the applications of machine learning fueled by vast new amounts of “big data” have grown exponentially. Advertisement placements, dating recommendations, genetic medicines, baseball, stock market investing, and self-driving cars are just a few domains that are heavily driven by machine learning technologies. Ethical issues are emerging from these developments that still have not been fully resolved by society. But one thing appears certain: machine learning is here to stay.

In this course you have read about many of the foundational concepts and techniques of machine learning. This task asks you create a detailed proposal concerning how machine learning techniques could be applied to a relevant data set of your choosing, what possible outcomes might be expected, and why such outcomes could be organizationally significant.

SCENARIO

In this assignment, you will assume the role of a recently hired machine learning engineer in an organization that has been asked to use available data to identify an organizational need that could be solved by machine learning. The organization then needs to outline a plan for designing and executing a machine learning model to solve this organizational need. You will explore available data sets and machine learning algorithms supporting the construction of a model that the organization would use, following the plan that you put



together. You will outline the problem, the solution, and the project plan, as well as the framework for evaluating the success of the model and the project.

REQUIREMENTS

Your submission must be your original work. No more than a combined total of 30% of the submission and no more than a 10% match to any one individual source can be directly quoted or closely paraphrased from sources, even if cited correctly. The originality report that is provided when you submit your task can be used as a guide.

You must use the rubric to direct the creation of your submission because it provides detailed criteria that will be used to evaluate your work. Each requirement below may be evaluated by more than one rubric aspect. The rubric aspect titles may contain hyperlinks to relevant portions of the course.

*Tasks may **not** be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc., unless specified in the task requirements. All other submissions must be file types that are uploaded and submitted as attachments (e.g., .docx, .pdf, .ppt).*

Project Overview

A. Create a proposal for a machine learning project by doing the following:

1. Describe an organizational need that your project proposes to solve.
2. Describe the context and background for your project.
3. Review **three** outside works that explore machine learning solutions that apply to the need described in part A1.

Note: These works may include interviews, white papers, research studies, or other types of work by industry professionals. Works that support your research may be identified from various sources, including the WGU Library.

- a. Describe how *each* reviewed work from part A3 relates to the development of your project.
4. Summarize the machine learning solution you plan to use to address the organizational need described in part A1.
5. Describe the benefits of your proposed machine learning

Machine Learning Project Design

B. Describe your proposed machine learning project plan by doing the following:

1. Define the scope of the proposed machine learning project.
2. Explain the goals, objectives, and deliverables for the proposed project.
3. Explain how you will apply a standard methodology (e.g., CRISP-DM, SEMMA) to the implementation of your proposed project.
4. Provide a projected timeline for the proposed project, including the start and end dates for *each* task.
5. List resources (e.g., hardware, software, work hours, third-party services) and *all* associated costs needed to implement the proposed solution.
6. Describe the criteria that you will use to evaluate the success of the project once it is completed.

Machine Learning Solution Design

C. Describe the proposed machine learning solution you will use to address the organizational need identified in part A1 by doing the following:

1. Identify the hypothesis of the proposed project.
2. Identify the machine learning algorithm(s) (i.e., supervised, unsupervised, or reinforcement learning) you will implement in your proposed solution.
 - a. Justify the selection of the algorithm in part C2. Include **one** advantage and **one** limitation of the selected machine learning method.
3. Describe the tools and environments that will be used to develop the proposed machine learning solution, including any third-party code.
4. Explain the process you will use to measure the performance of your proposed machine learning solution.

Description of Data Set(s)

- D. Describe the data for your proposed project by doing the following:
1. Identify the source(s) of the data for your proposed project.
 2. Describe the data collection method.
 - a. Discuss **one** advantage and **one** limitation of the data collection method described in part D2.
 3. Explain how you will prepare your data for use by the machine learning algorithm(s) from part C2 for your proposed project, including data set formatting, missing data, outliers, dirty data, or mitigation of other data anomalies.
 4. Describe behaviors that should be exercised when working with and communicating about sensitive data in your project.
- E. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.
- F. Demonstrate professional communication in the content and presentation of your submission.

File Restrictions

File name may contain only letters, numbers, spaces, and these symbols: ! - _ . * ' ()

File size limit: 400 MB

File types allowed: doc, docx, rtf, xls, xlsx, ppt, pptx, odt, pdf, txt, qt, mov, mpg, avi, mp3, wav, mp4, wma, flv, asf, mpeg, wmv, m4v, svg, tif, tiff, jpeg, jpg, gif, png, zip, rar, tar, 7z

RUBRIC

A1:ORGANIZATIONAL NEED

NOT EVIDENT

A description of an organizational need that the machine learning project proposes to solve is not provided.

APPROACHING COMPETENCE

The description of an organizational need that the machine learning project proposes to solve is incomplete, inaccurate, or unfocused.

COMPETENT

The description of an organizational need the machine learning project proposes to solve is complete, accurate, and focused.

A2:PROJECT CONTEXT AND BACKGROUND

NOT EVIDENT

A description of the context and background for the machine learning project is not provided.

APPROACHING COMPETENCE

The description of the context or background of the machine learning project is incomplete or inaccurate.

COMPETENT

The description of the context and background of the machine learning project is *both* complete and accurate.

A3: REVIEW OF BACKGROUND WORKS**NOT EVIDENT**

A review of *all* 3 outside works that explore the need described in part A1 is not provided.

APPROACHING COMPETENCE

The review of 1 or more outside works does not explore machine learning solutions, or 1 or more of the works do not apply to the need described in part A1. Or the review of the works is inaccurate or incomplete.

COMPETENT

The review of *all* 3 outside works accurately and completely explores machine learning solutions that apply to the need described in part A1.

A3A: RELATION TO PROJECT DEVELOPMENT**NOT EVIDENT**

A description of how *each* of the 3 reviewed works from part A3 relates to the development of the project is not provided.

APPROACHING COMPETENCE

The description of how 1 or more of the 3 reviewed works from part A3 relates to the development of the project is inaccurate or incomplete.

COMPETENT

The description of how *each* of the 3 reviewed works from part A3 relates to the development of the project is accurate and complete.

A4: SUMMARY OF THE MACHINE LEARNING SOLUTION**NOT EVIDENT**

A summary of the machine learning solution to address the organizational need described in part A1 is not provided.

APPROACHING COMPETENCE

The summary of the planned machine learning solution does not specifically align to the organizational need described in part A1, or the summary is inaccurate.

COMPETENT

The summary of the planned machine learning solution specifically and accurately aligns to the organizational need described in part A1.

A5: BENEFITS OF THE MACHINE LEARNING SOLUTION**NOT EVIDENT****APPROACHING COMPETENCE****COMPETENT**

A description of the benefits of the proposed machine learning solution is not provided.

The description of the benefits of the proposed machine learning solution is inaccurate or inadequate.

The description of the benefits of the proposed machine learning solution is accurate and adequate.

B1:SCOPE OF PROJECT**NOT EVIDENT**

A definition of the machine learning project's scope is not provided.

APPROACHING COMPETENCE

The definition of the machine learning project's scope is inaccurate or incomplete.

COMPETENT

The definition of the machine learning project's scope is *both* accurate and complete.

B2:GOALS, OBJECTIVES, AND DELIVERABLES**NOT EVIDENT**

An explanation of the proposed project's goals, objectives, and deliverables is not provided.

APPROACHING COMPETENCE

The explanation does not include the project's goals, objectives, or deliverables, or the explanation is incomplete or inaccurate.

COMPETENT

The explanation includes the proposed project's goals, objectives, and deliverables, and the explanation is complete and accurate.

B3:STANDARD METHODOLOGY**NOT EVIDENT**

An explanation of a standard methodology is not provided nor how it will be applied to the implementation of the proposed project.

APPROACHING COMPETENCE

The explanation does not accurately or completely explain how the standard methodology will be applied to the implementation of the proposed project.

COMPETENT

The explanation accurately and completely explains how the standard methodology will be applied to the implementation of the proposed project.

B4:TIMELINE AND MILESTONES**NOT EVIDENT**

A projected timeline with tasks is not provided.

APPROACHING COMPETENCE

The projected timeline for the proposed project is incomplete or does not include start and end dates for *each* task, or the task dates are inaccurate or illogical.

COMPETENT

The projected timeline for the proposed project is complete and includes accurate and logical start and end dates for *each* task.

B5:RESOURCES AND COSTS**NOT EVIDENT**

A list of resources and associated costs needed to implement the proposed solution is not provided.

APPROACHING COMPETENCE

The list does not include *all* resources or *all* associated costs needed to implement the proposed solution. Or the listed resources or costs are inaccurate.

COMPETENT

The list includes *all* resources and *all* estimated costs needed to implement the proposed solution. The listed resources and costs are accurate.

B6:CRITERIA FOR SUCCESSFUL EXECUTION OF PROJECT**NOT EVIDENT**

A description of criteria used to evaluate the success of the completed project is not provided.

APPROACHING COMPETENCE

The description of the criteria used to evaluate the success of the completed project is incomplete or inaccurate.

COMPETENT

The description of the criteria used to evaluate the success of the completed project is *both* complete and accurate.

C1:HYPOTHESIS**NOT EVIDENT**

A hypothesis for the proposed project is not identified.

APPROACHING COMPETENCE

The identified hypothesis for the proposed project is illogical or incomplete.

COMPETENT

The identified hypothesis for the proposed project is *both* logical and complete.

C2:ANALYTICAL METHOD**NOT EVIDENT**

A machine learning algorithm or algorithms that will be implemented in the proposed solution is not identified.

APPROACHING COMPETENCE

The machine learning algorithm or algorithms identified is inappropriate for the proposed project.

COMPETENT

The identified machine learning algorithm or algorithms is appropriate for the proposed project.

C2A:JUSTIFICATION OF ALGORITHM SELECTION**NOT EVIDENT**

A justification of the machine learning algorithm identified in part C2 is not provided.

APPROACHING COMPETENCE

The justification of the machine learning algorithm identified in

COMPETENT

The justification of the machine learning algorithm identified in part C2 is logical, accurate, and

part C2 is illogical, inaccurate, or incomplete. Or it does not include 1 advantage or 1 limitation of the selected algorithm.

complete. And it includes 1 advantage and 1 limitation of the selected algorithm.

C3:TOOLS AND ENVIRONMENTS OF SOLUTION**NOT EVIDENT**

A description of the tools and environments that will be used to develop the proposed machine learning solution is not provided.

APPROACHING COMPETENCE

The description of the tools or the environments that will be used to develop the proposed machine learning solution is inaccurate, or it does not identify any third-party code that will be used.

COMPETENT

The description accurately describes the tools and the environments that will be used to develop the proposed machine learning solution, and it identifies any third-party code that will be used.

C4:MEASURING PERFORMANCE**NOT EVIDENT**

An explanation of the process used to measure the performance of the machine learning solution is not provided.

APPROACHING COMPETENCE

The explanation of the process used to measure the performance of the machine learning solution is inaccurate or incomplete.

COMPETENT

The explanation of the process used to measure the performance of the machine learning solution is *both* accurate and complete.

D1:SOURCE OF DATA**NOT EVIDENT**

The source or sources for the data used in the proposed project are not identified.

APPROACHING COMPETENCE

The source or sources for the data used in the proposed project are inaccurate or incomplete.

COMPETENT

The source or sources for the data used in the proposed project are *both* accurate and complete.

D2:DATA COLLECTION METHOD**NOT EVIDENT**

A description of the data collection method is not provided.

APPROACHING COMPETENCE

The description of the data collection method is incomplete or inaccurate.

COMPETENT

The description of the data collection method is *both* complete and accurate.

D2A:ADVANTAGES AND LIMITATIONS**NOT EVIDENT**

A discussion of 1 advantage and 1 limitation of the data collection method described in part D2 is not provided.

APPROACHING COMPETENCE

A discussion of *both* 1 advantage and 1 limitation of the data collection method described in part D2 is provided, but the discussion of 1 or *both* is inaccurate or incomplete.

COMPETENT

The discussion of *both* 1 advantage and 1 limitation of the data collection method described in part D2 is *both* accurate and complete.

D3:QUALITY AND COMPLETENESS OF DATA**NOT EVIDENT**

An explanation of how the data will be prepared for use by the machine learning algorithm or algorithms from part C2 for the proposed project is not provided.

APPROACHING COMPETENCE

The provided explanation of how the data will be prepared for use by the machine learning algorithm or algorithms from part C2 for the proposed project is incomplete or incorrect.

COMPETENT

The provided explanation of how the data will be prepared for use by the machine learning algorithm or algorithms from part C2 for the proposed project is *both* complete and correct.

D4:PRECAUTIONS FOR SENSITIVE DATA**NOT EVIDENT**

The description of behaviors that should be exercised when working with and communicating about sensitive data in the project is not provided.

APPROACHING COMPETENCE

The description of the behaviors that should be exercised when working with or communicating about sensitive data is inaccurate or illogical.

COMPETENT

The description of the behaviors that should be exercised when working with and communicating about sensitive data in the project is *both* accurate and logical.

E:SOURCES**NOT EVIDENT**

The submission does not include both in-text citations and a reference list for sources that are quoted, paraphrased, or summarized.

APPROACHING COMPETENCE

The submission includes in-text citations for sources that are quoted, paraphrased, or summarized and a reference list; however, the citations or reference list is incomplete or inaccurate.

COMPETENT

The submission includes in-text citations for sources that are properly quoted, paraphrased, or summarized and a reference list that accurately identifies the author, date, title, and source location as available.

F:PROFESSIONAL COMMUNICATION

NOT EVIDENT

Content is unstructured, is disjointed, or contains pervasive errors in mechanics, usage, or grammar. Vocabulary or tone is unprofessional or distracts from the topic.

APPROACHING COMPETENCE

Content is poorly organized, is difficult to follow, or contains errors in mechanics, usage, or grammar that cause confusion. Terminology is misused or ineffective.

COMPETENT

Content reflects attention to detail, is organized, and focuses on the main ideas as prescribed in the task or chosen by the candidate. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar promote accurate interpretation and understanding.

REFERENCE LIST

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and Tensorflow* (2nd ed.). O'Reilly Media.