Instructions for Task-3

C951 – Introduction to Artificial Intelligence (AI)

by

Robert Ferdinand (CS Instructor @ WGU)
robert.ferdinand@wgu.edu

Initial Recommendations

1. For your paper, use the provided TEMPLATE.

2. For Industry Examples of Machine Learning CLICK and CLICK.

A: Project Overview

1. Describe a need of an organization that your project will solve: Provide concrete details of the organization's need that your project proposes to address. What problem currently exists in the organization and how will your Machine Learning (ML) solution solve it!

2. Describe the context and background of your project: Include some history of the organization and its industry. How will the ML solution benefit the organization? Provide suitable references with sources listed in the last section of your submission.

A: Project Overview (Continued)

3. Fully review three external works (interviews, research papers etc.) that explore ML solutions applicable to the need in A.1. Describe in detail how each review relates to the development of your project. Source: *scholar.google.com*. Make sure to cite the sources!

4. Summarize how you plan to use your ML solution to address the need in A.1. Describe your ML model here. What kind of ML model do you propose? Supervised, Unsupervised, Recommendation, Association, Deep Learning. Can you be even more specific?

A: Project Overview (Continued)

5. Discuss the benefits of your ML solution. How will your ML solution help the organization? How do you justify its application to the need in A.1.

B: Machine Learning (ML) Project Design

1. Define scope of the proposed ML project:
   a) Outline project's goals, deliverables (end products), tasks, project members, deadlines and milestones (progress markers) (Make a List).
   b) List items that are "Out of Scope" for project (Make a List).

2. Explain project goals, objectives, and deliverables in detail. Use three subsections:
   a) Goals: Productivity Goals, Cost Goals, Improvement Goals.
   b) Objectives: Specific, Quantifiable (numeric). Examples: Financial, % of Uptime (Availability), Accuracy, Throughput (Rate). Note: Also stated in B.6.

B: Machine Learning (ML) Project Design (Continued)

c) Deliverables: ML model, Hardware/Software, Data Sources.

3. Apply a standard methodology CRISP-DM or SEMMA to your project. *(Discuss each phase and provide concrete example for each)*

4. Provide a projected Timeline for your ML project: Split project into tasks with start and end dates for each task. Task Examples: Obtaining Physical Workspace, Creating AI Framework, Networking, Training Data, Data Pipeline Creation and Training, Testing Data, Exporting Pipeline, Cloud Availability, Project Approval by Sponsor, Project Implementation, Creating GUI, Final Project Testing, Connecting to Database, Team Account Setup, Final Project Signoff.

B: Machine Learning (ML) Project Design (Continued)

5. Resources required for ML model and Costs for each: Includes hardware, software, cloud hosting, data, team development (programmers, specialists, designers etc.). Make a table. Here is a VIDEO.

6. Describe criteria used to evaluate project's success upon completion. Evaluate objectives from 2 b) above. Criteria include accuracy, throughput, stability, financial, model success etc. Make a table.

C: Machine Learning (ML) Solution Design

1. Identify project hypothesis. What is the current need? How will your ML model address the need? What methods will your ML model use to address the need? Here is a VIDEO on hypothesis.

2. What ML algorithm will you use? (Supervised, Unsupervised, Reinforcement Learning). You can get even more specific. Justify why this algorithm. Include one advantage (speed, simplicity, accuracy, appropriateness etc.) and one disadvantage (overfitting, etc.) of your algorithm.

3. Describe tools such as programming language, environment, operating system, libraries you will use for your ML model.

C: Machine Learning (ML) Solution Design (Continued)

4. How will you measure performance of your ML model? Talk about:
   a) Quantity Measured such as Financial Profit, Rate Of Interest, Efficiency, Correctness, Customer or Employee Satisfaction, Time or Cycle Time Length, Productivity, Throughput, Error Rates etc.
   b) Training and Testing Data Sets Used in Modeling.
   c) Metrics for Measuring Accuracy.
   d) Maximum Allowed Error.

D: Description of Data Sets

1. Identify Source of Data: May not be an actual, live, accessible data source unless you plan to use it for your Capstone project. Describe the origins of the Data set.

2. Describe Your Data Collection:
   a) From where and how you will acquire your Data set for the ML system?
   b) List and Fully Explain one Advantage and one Limitation of the Data Collection method from 2 a).

D: Description of Data Sets (Continued)

3. Preparing Data for ML model from C.2. Consider:

a) Formatting Data Set.
b) Missing Data, Outliers, Dirty Data, Null Values, Anomalies.
c) Time origin of data for relevance.
d) ETL (Extract, Transform, Load) for Data.
e) Cleaning Data of PII (Personally Identifiable Information).
f) Relevance of all Data Fields in the Data set.
g) Uniformity between Yes/No, True/False, On/Off Boolean variables.
h) Keeping Data current – Updating regularly.

D: Description of Data Sets (Continued)

4. Describe behaviors to be exercised while working with and communicating about large sets of sensitive data. Discuss the following:

   a) Security and Risk of Theft.
   b) Loss of Data.
   c) Corruption of Data.
   d) Internal Theft (by Employees).
   e) Non-compete Agreements.

   * If there is NO sensitive data mention that here!

E: Acknowledge sources (papers, interviews, books, articles etc.), using in-text citations and references, for content that is quoted, paraphrased or summarized.

F: Ensure that your submission is professional in its content and presentation.