

# DAT 640 Final Project Guidelines and Rubric

## Overview

The final project for this course is the creation of a predictive data analytic model with documentation and visualizations that will serve as a report for stakeholders. Your job is to accurately communicate the potential value that a predictive model could add to an organization, using your selected data source.

The project is divided into **five milestones**, which will be submitted at various points throughout the course to scaffold learning and ensure quality final submissions. These milestones will be submitted in **Modules Two, Three, Five, Seven, and Eight**. The final submission will occur in **Module Nine**.

This project will assess your mastery with respect to the following **course outcomes**:

- Recommend appropriate types of predictive algorithms for use in data analysis scenarios with selective structure, interaction, and relationship specifications
- Select appropriate tools for successful predictive algorithm implementation within various environments
- Evaluate predictive model structure, interactions, and relationships in data for reliability using common scoring models
- Propose valid and measurable methods for continuous predictive modeling improvement by employing guiding scoring models and implementing continuous feedback
- Defend predictive model design and pilot implementation plans by communicating the value of predictive analytics within the context of specific scenarios

## Prompt

In order to produce your analytic report, you will need to choose and investigate a data set using the data mining techniques you learned in class. Then you will formulate a research question, make a plan, and implement it. Your report should include detailed descriptions of the process, the explorations of the meaning behind your model, and the implications of its results. You will also be testing your model's fitness and evaluating its strengths and weaknesses. The key components of the project are as follows:

1. Perform a basic summary analysis.
2. Formulate your research question.
3. Write an analytic plan.
4. Perform predictive analytics.
5. Perform model diagnostics.
6. Evaluate.

Your report should answer the following question: What potential value can be added to your selected organization by employing a predictive analytic model, and how would you implement such a model?

A predictive analytics/data mining report is similar to any other analytic report. One key difference is that the “research question” is prospective, rather than hypothesis answering. For example, a data mining research question might be to find the best model for a process, using a large data set. Or a predictive model might seek the most useful set of predictors for a given known outcome. Or a different predictive model might seek the threshold values at which the model changes from one recommendation to another, and leave it to the reader to decide what to do. That is the nature of prognostication: It is open to interpretation. Your job is to make the scenario clear.

Predictive modeling seeks less to explain the future and more to explain the possibilities that may lead to various futures. This is the modeling aspect. If you model the weather but never take into account barometric pressure, your model will fail if you try to decide hurricane trajectories. So, these are the kinds of things you will be looking at in your predictive models: searching for ways to explain the conditions that produce outcomes and to evaluate the strengths and weaknesses of the models you produce.

The three main ideas that your report should encompass are your ability to formulate a prospective research question, to assess and model data, and, finally, to evaluate the relative strengths and weaknesses of a model. In short, if your report fully encompasses these three ideas, you will produce an authentic document that would stand on its own in a professional setting.

#### Data Source

Your final project will be based on real-world data that was prepared specifically for a Computational Intelligence and Learning (Coll) Challenge held in 2000. Information about the challenge and a link to the data [can be found at this website](#). Instructions, if you need them, on how to download the data into the VDI will be provided separately. Ground rules for using this data are as follows:

- Use the data set TICDATA2000.txt for all the activities related to this final project. Note: This data is a tab-delimited file with no column headings. We will provide a tutorial on how to assign column heading names separately, if you need that guidance.
- TICEVAL2000.txt is a validation data set with additional unscored records. **Do not use this data until you are ready to test your final model.**
- The scores for these records are held out in another file, TICTGTS2000.txt. The results from applying your model blindly to this unscored data set will give you the best feedback on how well your model captured the relationships between the input variables and the outcomes. Submit a summary of your results against the validation data set with your final submission.
- There are some references regarding this contest and the entries and commentary on what worked and what did not. Your personal learning experience will be maximized if you spend time, up-front, thinking about your own approach and trying different methods for improving your model **before** you read any of the literature. Provide citations for any published work that you read in support of completing this project.

Specifically, you must address the following **critical elements** in your final submission:

- I. **Organizational Background**
  - A. Provide a **general background** of the organization you have selected (industry, size, etc.). What are the problems or limitations the organization hopes to overcome?
  - B. What is the **potential value** that developing a predictive analytic strategy holds for the organization? In other words, how would development of a predictive analytic strategy help the organization?
  
- II. **Predictive Algorithms**
  - A. **What specifications** will you need to follow in selecting a predictive algorithm? Specifically, explain the selective structure, interaction, and relationship specifications of your organization.
  - B. Given the specifications put forth by the business, what types of predictive algorithms may be appropriate for use?
    - i. **Recommend** a predictive algorithm that would be appropriate for the selective structure, interaction, and relationship specifications of your chosen organizational data.
  
- III. **Data Analytic Tools**
  - What tools would facilitate implementation of your chosen predictive algorithm? Evaluate each tool and describe how it would support implementation of the chosen predictive algorithm.
  
- IV. **Model Optimization**
  - A. Analyze the **commonly used approaches and methods** of scoring predictive models, and determine what method you would use. Why have you selected this method?
  - B. **Evaluate** the predictive model structure, interactions, and relationships of your chosen scoring method for reliability. Specifically, how reliable would data be as it moves through the predictive model structure you are proposing?
  - C. Describe the **steps for implementing** a continuous feedback method “scoring engine” for measuring the validity and improvement of a predictive analytic model.
  - D. Document the **expected results** of using the scoring engine for optimization of the predictive analytic model.
  
- V. **Reproducible Research**
  - A. Document your predictive model design using reproducible research **techniques** so that others may reproduce the design in various contexts.
  - B. Propose a **pilot plan** of predictive analytic strategy that is consistent with the organization specifications.
  - C. Create a brief **presentation** that communicates the value of the predictive analytics within the context of your selected organization for governance.

## Milestones

### Milestone One: Describe a Data Set for the Final Project

In **Module Two**, you will describe the data set for the final project. You will provide a general background on the industry, organization, a description of the specific issue facing the organization, and a basic description of the data set. **This milestone will be graded with the Milestone One Rubric.**

### Milestone Two: Introduction

In **Module Three**, you will develop the predictive analytic question and analytic plan. You will also describe how that strategy will help solve a specific issue within the organization and try to characterize the specific value that will be achieved. **This milestone will be graded with the Milestone Two Rubric.**

### Milestone Three: Problem Statement/Research

In **Module Five**, you will develop your problem statement or research question. **This milestone will be graded with the Milestone Three Rubric.**

### Milestone Four: Pilot Plan

In **Module Seven**, you will create a pilot plan using reproducible research design techniques. **This milestone will be graded with the Milestone Four Rubric.**

### Milestone Five: Model Optimization

In **Module Eight**, you will address model optimization, including an evaluation of the predictive model, steps for implementing a continuous feedback method for measuring the validity of the model, and the expected results of using the scoring engine for optimization. **This milestone will be graded with the Milestone Five Rubric.**

### Final Project Submission: Predictive Data Analytic Model

In **Module Nine**, you will submit your final project. It should be a complete, polished artifact containing all of the critical elements of the final product. It should reflect the incorporation of feedback gained throughout the course. **This submission will be graded with the Final Project Rubric.**

## Deliverables

Milestone	Deliverable	Module Due	Grading
One	Describe a Data Set for the Final Project	Two	Graded separately; Milestone One Rubric
Two	Introduction	Three	Graded separately; Milestone Two Rubric
Three	Problem Statement/Research	Five	Graded separately; Milestone Three Rubric
Four	Pilot Plan	Seven	Graded separately; Milestone Four Rubric
Five	Model Optimization	Eight	Graded separately; Milestone Five Rubric
	Final Submission: Predictive Data Analytic Model	Nine	Graded separately; Final Project Rubric

## Final Project Rubric

**Guidelines for Submission:** Your paper should be 8 to 10 pages in length (plus a title page and references). It should be submitted as a Microsoft Word document with double spacing, 12-point Times New Roman font, and one-inch margins. Your citations should follow APA style.

Critical Elements	Exemplary (100%)	Proficient (90%)	Needs Improvement (70%)	Not Evident (0%)	Value
<b>Organizational Background: General</b>	Meets "Proficient" criteria, and the problems are contextualized around what could be addressed by implementation of a predictive analytics structure	Provides an accurate background of the organization and describes the problems/limitations that the organization wants to overcome	Provides a background of the organization, but there is a lack of specificity in the description of the problems/limitations that the organization wants to overcome	Does not provide a background of the organization	5
<b>Organizational Background: Potential Value</b>	Meets "Proficient" criteria and added value is articulated persuasively for multiple audiences	Logically explains the possible value that could be added to the organization by the development of a predictive analytic strategy	Explanation of the possible value that could be added to the organization by the development of a predictive analytic strategy is not logical	Does not explain the possible value that could be added to the organization by the development of a predictive analytic strategy	5
<b>Predictive Algorithms: Specifications</b>	Meets "Proficient" criteria, and explanations are qualified with real-world examples for support	Accurately determines and explains the selective structure, interaction, and relationship specifications of the organization	Determines and explains the selective structure, interaction, and relationship specifications of the organization, but explanations or selections are not always accurate	Does not determine or explain the selective structure, interaction, and relationship specifications of the organization	10

<b>Predictive Algorithms: Recommendation</b>	Meets "Proficient" criteria and cites relevant, professional resources to support the defense	Recommends a predictive algorithm and accurately defends its utility in meeting the specifications of the organization	Recommends the predictive algorithm but does not accurately defend its utility in meeting the specifications of the organization	Does not recommend or defend the utility of the predictive algorithm in meeting organizational specifications	5
<b>Data Analytic Tools</b>	Meets "Proficient" criteria and evaluation is accurately contextualized around implementation within the organization	Selects and accurately evaluates the utility of tools that would support implementation of the predictive algorithm	Selects and analyzes the utility of tools that would support implementation of the predictive algorithm, but is not always accurate	Does not select and analyze the utility of tools that would support implementation	5
<b>Model Optimization: Commonly Used Approaches and Methods</b>	Meets "Proficient" criteria, and analysis is qualified with valid and scholarly citations	Analyzes the commonly used approaches and methods of scoring predictive models in detail within the context of the organization	Analyzes the commonly used approaches and methods of scoring predictive models, but there is a lack of detail in the context of the organization	Does not analyze the commonly used approaches and methods of scoring predictive models	10
<b>Model Optimization: Evaluation</b>	Meets "Proficient" criteria, and supporting research is scholarly in nature	Evaluation of the predictive model structure, interactions, and relationships in data of the selected scoring method for reliability is logical and supported with relevant research	Evaluation of the predictive model structure, interactions, and relationships in data of the selected scoring method for reliability is not logical or is not supported with relevant research	Does not include an evaluation of the predictive model structure, interactions, and relationships in data of the selected scoring method for reliability	10
<b>Model Optimization: Implementation Steps</b>	Meets "Proficient" criteria, and plan is comprehensive enough to be put into action within the context of the organization	Accurately describes the necessary steps for implementing a continuous feedback method "scoring engine" for measuring the validity and improvement of a predictive analytic model in detail	Describes the steps for implementing a continuous feedback method "scoring engine" for measuring the validity and improvement of a predictive analytic model, but misses necessary detail or does not accurately identify the steps	Does not describe the steps for implementing a continuous feedback method "scoring engine" for measuring the validity and improvement of a predictive analytic model	10
<b>Model Optimization: Expected Results</b>	Meets "Proficient" criteria, and explanation is diversified to include possible unexpected results	Documents and clearly explains the expected results of using the scoring engine for optimization of the predictive analytic model	Documentation and explanation of the expected results of using the scoring engine for optimization of the predictive analytic model are not clear	Does not document or explain the expected results of using the scoring engine for optimization of the predictive analytic model	5
<b>Reproducible Research: Techniques</b>	Meets "Proficient" criteria, and documentation is qualified with brief narrations around choices for external audiences	Clearly and accurately documents the design using reproducible research techniques so that others may	Documents the design using reproducible research techniques, but documentation lacks clarity or specificity	Does not document the design using reproducible research techniques	5

	reproduce the design in other contexts	necessary so that others may reproduce the design		
<b>Reproducible Research: Pilot Plan</b>	Meets "Proficient" criteria and cites specific examples to show how the pilot will test against the organization specifications	Proposes a reasonable pilot plan for the predictive analytic strategy within the context of the organization specifications	Proposes a pilot plan for the predictive analytic strategy, but the plan is not entirely reasonable given the context of the organization specifications	Does not propose a pilot plan for the predictive analytic strategy
<b>Reproducible Research: Presentation</b>	Meets "Proficient" criteria, and presentation is articulated in a persuasive manner supported with scholarly materials	Creates a brief presentation that effectively articulates the value of the predictive analytic strategy for governance of the organization	Creates a brief presentation that explains the value of the predictive analytic strategy for governance of the organization, but the articulation is not always effective	Does not create a brief presentation that explains the value of the predictive analytic strategy for governance
<b>Articulation of Response</b>	Submission is free of errors related to citations, grammar, spelling, syntax, and organization and is presented in a professional and easy-to-read format	Submission has no major errors related to citations, grammar, spelling, syntax, or organization	Submission has major errors related to citations, grammar, spelling, syntax, or organization that negatively impact readability and articulation of main ideas	Submission has critical errors related to citations, grammar, spelling, syntax, or organization that prevent understanding of ideas
				<b>Earned Total   100%</b>