|  |
| --- |
| **Assurance of Student Learning Report****2021-2022** |
| *Ogden College of Science and Engineering* | *School of Engineering and Applied Sciences* |
| *Scientific Data Analytics certificate 0496* |
| *Zhonghang Xia* |

|  |
| --- |
| ***Use this page to list learning outcomes, measurements, and summarize results for your program. Detailed information must be completed in the subsequent pages.*** |
| **Student Learning Outcome 1: Write computer programs to utilize and analyze large datasets.**  |
| **Instrument 1** | **Assignment 1 in CS 555** |
| **Instrument 2** | **Assignment 3 in CS 555** |
| **Instrument 3** |  |
| **Based on your results, check whether the program met the goal Student Learning Outcome 1.** | **[x]  Met** | **[ ]  Not Met** |
| **Student Learning Outcome 2:** **Understand the statistical approaches taken when dealing with large sample sizes.** |
| **Instrument 1** | **Final exam in CS 555** |
| **Instrument 2** |  |
| **Based on your results, check whether the program met the goal Student Learning Outcome 2.** | **[x]  Met** | **[ ]  Not Met** |
|  |  |  |
| **Student Learning Outcome 3: Understand the statistical approaches taken when dealing with multiple variables.**  |
| **Instrument 1** | **Term project in CS 555** |
| **Instrument 2** |  |
| **Based on your results, check whether the program met the goal Student Learning Outcome 3.** | **[x]  Met** | **[ ]  Not Met** |
| **Student Learning Outcome 4: Combine domain expertise with programming and statistical skills to analyze large domain‐specific datasets.** |
| **Instrument 1** | **Term project in CS 555** |
| **Instrument 2** |  |
| **Based on your results, check whether the program met the goal Student Learning Outcome 4.** | **[x]  Met** | **[ ]  Not Met** |
|  |  |  |
| **Program Summary (Briefly summarize the action and follow up items from your detailed responses on subsequent pages.)**  |
| Overall, the results from this assessment indicate that the program has reached and exceeded the self-reported assessment goals in each category. More explicit guidelines for assessing the skills of analyzing domain-specific datasets will be established when CS 555 is offered in Fall 2022.  |

|  |
| --- |
| **Student Learning Outcome 1** |
| **Student Learning Outcome**  | **Write computer programs to utilize and analyze large datasets.** |
| **Measurement Instrument 1**  | Students are required to use Python or other programs to retrieve data from large-sized data files, manipulate the data for data analytics, and output analytic results.  |
| **Criteria for Student Success** | Students should reach at least the intermediate level of performance.  |
| **Program Success Target for this Measurement** | **80%** | **Percent of Program Achieving Target** | **80%** |
| **Methods**  | Students are required to write python programs to read users’ ratings on movies from four moive rating files. The next step is data cleaning. Students need to check whether there exist any missing data and duplication before analysis. The cleaned data samples are saved in the format of pandas dataframe or numpy array according. Students use builit-in pandas functions to count total numbers of users and items, and then compute the average ratings of specific group of users. the student needs to select some rows and columns of the dataset for the required features. Finally, students are required to write python programs to display the analytic results on the screen and also save the results into files. Rubrics for O1 are used.  |
| **Measurement Instrument 2** | Students are required to use Python built-in methods to obtain analytic results.  |
| **Criteria for Student Success** | Students should reach at least the intermediate level of performance.  |
| **Program Success Target for this Measurement** | **80%** | **Percent of Program Achieving Target** | **80%** |
| **Methods** | Students needs to select correct analytical model for predicting the levels of air pollution. Linear regression and LASSO are two commonly used machine learning models to describe linear relationship in the dataset. Students use part of the air pollution dataset to train the models and then predict the levels of PM2.5 with the models on the other part of the dataset. The prediction results are evaluated by metrics R-square, Mean Square Error, and Root of Mean Square Error. Rubric item O1-2 are used. |
| **Based on your results, highlight whether the program met the goal Student Learning Outcome 1.** | **[x]  Met** | **[ ]  Not Met** |
| **Actions** (Describe the decision-making process and actions for program improvement. The actions should include a timeline.) |
| Python program was used in CS 555 in Spring 2022 although other programming languages can also be used in this course. Data input, output, and pre-processing were examined in relation to the certificate learning outcome. No change is needed.  |
| **Follow-Up** (Provide your timeline for follow-up. If follow-up has occurred, describe how the actions above have resulted in program improvement.) |
| The rubrics were first used in Spring 2021 and will be followed next time.  |
| **Next Assessment Cycle Plan** (Please describe your assessment plan timetable for this outcome) |
| This outcome will be assessed when CS 555 is offered next time. According to the current course rotation schedule of the Computer Science Program, CS 555 will be offered in Spring 2023. |

|  |
| --- |
| **Student Learning Outcome 2** |
| **Student Learning Outcome**  | **Understand the statistical approaches taken when dealing with large sample sizes.** |
| **Measurement Instrument 1** | **Build a relationship between independent and dependent variables with the logistic regression model. Evaluate the selected model.** |
| **Criteria for Student Success** | Students should reach at least the intermediate level of performance.  |
| **Program Success Target for this Measurement** | **80%** | **Percent of Program Achieving Target** | **80%** |
| **Methods**  | A dataset of credit application is provided. The attribute names and values of the file have been changed to meaningless symbols to protect confidentiality of the data. There are 15 attributes plus a class label (+/-). Students are required to build a logistic regression to predict applicants’ credit status (+/-). We used Python 3 envirnment. A student’s program was first checked if there is any syntax error. If no syntax error, a test file would be fed into the program. The output of the test file was examed and compared with correct answers. Students are required to discuss the the performance of the logistic regression in terms of accuracy, precision and recall. Rubric item O2 was used. |
| **Based on your results, circle or highlight whether the program met the goal Student Learning Outcome 2.** | **[x]  Met** | **[ ]  Not Met** |
| **Actions** (Describe the decision-making process and actions planned for program improvement. The actions should include a timeline.) |
| No change is needed.  |
| **Follow-Up** (Provide your timeline for follow-up. If follow-up has occurred, describe how the actions above have resulted in program improvement.) |
| The rubrics were first used in Spring 2021 and will be followed next time.  |
| **Next Assessment Cycle Plan** (Please describe your assessment plan timetable for this outcome) |
| This outcome will be assessed when CS 555 is offered next time. According to the current course rotation schedule of the Computer Science Program, CS 555 will be offered in Spring 2023. |

|  |
| --- |
| **Student Learning Outcome 3** |
|  |
| **Student Learning Outcome**  | **Understand the statistical approaches taken when dealing with multiple variables.**  |
| **Measurement Instrument 1** | **Build a relationship between independent and dependent variables with the tree-based model. Evaluate the selected model.**  |
| **Criteria for Student Success** | Students should reach at least the intermediate level of performance.  |
| **Program Success Target for this Measurement** | **80%** | **Percent of Program Achieving Target** | **80%** |
| **Methods**  | A dataset of air pollution is provided, and the task is to predict predict the level of concentration of particle pollution.Students are required to build a tree-based machine learning models using PM10, CO, sulfur oxides, nitrogen oxides to predict the level of PM2.5. Students are allowed to use open-source libraries, such as scikit-learn. In the report, students discuss the the performance of the selected method and display their results in plots. Rubric item O3 are used. |
| **Based on your results, circle or highlight whether the program met the goal Student Learning Outcome 3.** | **[x]  Met** | **[ ]  Not Met** |
| **Actions** (Describe the decision-making process and actions for program improvement. The actions should include a timeline.) |
| No change is needed.  |
| **Follow-Up** (Provide your timeline for follow-up. If follow-up has occurred, describe how the actions above have resulted in program improvement.) |
| The rubrics were first used in Spring 2021 and will be followed next time.  |
| **Next Assessment Cycle Plan** (Please describe your assessment plan timetable for this outcome) |
| This outcome will be assessed when CS 555 is offered next time. According to the current course rotation schedule of the Computer Science Program, CS 555 will be offered in Spring 2023.  |

|  |
| --- |
| **Student Learning Outcome 4** |
|  |
| **Student Learning Outcome**  | **Combine domain expertise with programming and statistical skills to analyze large domain‐specific datasets.** |
| **Measurement Instrument 1** | Compare two machine learning algorithms in terms of algorithmic performance. |
| **Criteria for Student Success** | Students should reach at least the intermediate level of performance.  |
| **Program Success Target for this Measurement** | **80%** | **Percent of Program Achieving Target** | **80%** |
| **Methods**  | Students are required to build at least two different machine learning models on the dataset of air pollution. All models are implemented in Python 3. The performance of these models are compared in terms of MAE (Mean Absolute Error ), RMSE (Root Mean Square Error), ROC (Receiver operating characteristic). Students need to interpret the contribution of each feature to the prediction with partial dependence plot and SHAP. Rubric item O4 was used. |
| **Based on your results, circle or highlight whether the program met the goal Student Learning Outcome 3.** | **[x]  Met** | **[ ]  Not Met** |
| **Actions** (Describe the decision-making process and actions for program improvement. The actions should include a timeline.) |
| No change is needed. |
| **Follow-Up** (Provide your timeline for follow-up. If follow-up has occurred, describe how the actions above have resulted in program improvement.) |
| New modles were used in the evaluation in Spring 2022.  |
| **Next Assessment Cycle Plan** (Please describe your assessment plan timetable for this outcome) |
| This outcome will be assessed when CS 555 is offered next time. According to the current course rotation schedule of the Computer Science Program, CS 555 will be offered in Spring 2023. |

**Assessment Rubrics for Student Learning Outcome 1**

**Learning outcome 1:** **Write computer programs to utilize and analyze large datasets.**

Material Assessed: Assignment 1 and 3 in CS 555

Notes:

|  |  |  |  |
| --- | --- | --- | --- |
| **Evaluation item** | **Novice** | **Intermediate** | **Proficient** |
| **O1-1: Use Python or other programs to retrieve data from large-sized data files and manipulate the data for data analytics**  | Student cannot use a program to read specific data samples correctly. | Students can use a program to read specific data samples correctly and prepare the data in some formats. | Students can use a program to read specific data samples correctly and prepare the data in any desired format. |
| Number of students | 1 |  | 5 |
| **O1-2: Use Python or other programs to output analytic results** | Student cannot use a program to output analytic results correctly. | Student can use a program to output analytic results in some formats but not in a required format.  | Student can use a program to output analytic results in a required format. |
| Number of students | 1 |  | 5 |
| **O1-3: Use Python built-in methods to obtain analytic results** | Student cannot call correct methods to obtain analytic results. | Student knows which methods to use but cannot manage parameters correctly.  | Student can use right built-in methods and derive analytic results by choosing proper parameters in the methods |
| Number of students | 1 |  | 5 |

**Assessment Rubrics for Student Learning Outcome 2**

**Learning outcome 2:** **Understand the statistical approaches taken when dealing with large sample sizes.**

Material Assessed: Final exam in CS 555

Build a logistic model on the selected dataset, and evaluate the model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Evaluation item** | **Novice** | **Intermediate** | **Proficient** |
| **O2: Build a relationship between independent and dependent variables with the logistic regression model. Evaluate the selected model.**  | Student cannot build a model between dependent and independent model correctly. | Student can build a model but cannot analyze given datasets with the model correctly | Student can build a model and analyze given datasets with the model correctly.  |
| Number of students | 1 |  | 5 |

**Assessment Rubrics for Student Learning Outcome 3**

**Learning outcome 3:** **Understand the statistical approaches taken when dealing with multiple variables.**

Material Assessed: Part of the term project in CS 555

Build a tree-based model on the selected dataset and evaluate the model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Evaluation item** | **Novice** | **Intermediate** | **Proficient** |
| **O3: Build a relationship between independent and dependent variables with a tree-based model. Evaluate the selected model.**  | Student cannot build a multiple variable model correctly. | Student can build a multiple variable model but cannot analyze given datasets with the model correctly | Student can build a multiple variable model and analyze given datasets with the model correctly.  |
| Number of students | 1 |  | 5 |

**Assessment Rubrics for Student Learning Outcome 4**

**Learning outcome 4:** **Combine domain expertise with programming and statistical skills to analyze large**

**domain‐specific datasets.**

Material Assessed: Part of the term project in CS 555

Model the selected dataset with two different models and compare two machine learning algorithms in terms of algorithmic performance (MAE, RMSE, ROC) and interpretability.

|  |  |  |  |
| --- | --- | --- | --- |
| **Evaluation item** | **Novice** | **Intermediate** | **Proficient** |
| **O4: Compare two machine learning algorithms in terms of algorithmic performance**  | Student cannot build two machine learning models correctly. | Student can build models but cannot analyze the models correctly | Student can build the models and analyze the models correctly.  |
| Number of students |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CURRICULUM MAP** |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **Program name:** | Scientific Data Analytics, Certificate (0496) |   |  |  |  |
| **Department:** | School of Engineering and Applied Science |   |  |  |  |
| **College:** | Ogden College of Science and Engineering  |   |  |  |  |
| **Contact person:** | Zhonghang Xia |   |  |  |  |
| **Email:** | zhonghang.xia@wku.edu |   |  |  |  |
|   |  |  |  |  |  |  |
| **KEY:** |  |  |  |  |  |  |
| **I = Introduced** |  |  |  |  |  |  |
| **R = Reinforced/Developed** |  |  |  |  |  |
| **M = Mastered** |  |  |  |  |  |  |
| **A = Assessed** |  |  |  |  |  |  |
|  |  |  | **Learning Outcomes** |  |  |  |
|  |  |  | **LO1:** | **LO2:** | **LO3:** | **LO4:** |
|   |  |  | Write computer programs to utilize and analyze large datasets. | Understand the statistical approaches taken when dealing with large sample sizes | Understand the statistical approaches taken when dealing with multiple variables | Combine domain expertise with programming and statistical skills to analyze largedomain‐specific datasets |
| **Course Subject/Core** | **Number** | **Course Title** |   |   |   |   |
| STAT | 549 | Statistical methods |   | I/R/M | I/R/M |   |
| CS | 555 | Data science | I/R/M/A | A | A | I/R/M/A |