ACE3 — ACE3 TASK 1: DATA MINING REPORT

DATA MINING AND ANALYTICS II — C744

PRFA — ACE3

TASK OVERVIEW   SUBMISSIONS   EVALUATION REPORT

COMPETENCIES

4030.1.6 : **Review of Data Mining Process**
The graduate describes the methods and software for a data mining project.

4030.1.7 : **Descriptive Methods**
The graduate implements descriptive data mining methods.

4030.1.8 : **Classification and Prediction Methods**
The graduate implements classification and prediction data mining methods.

4030.1.9 : **Model Performance and Application**
The graduate assesses data mining model performance and application.

INTRODUCTION

One of the most critical factors in customer relationship management that directly impacts a company's long-term profitability is customer attrition. When a company can better predict if a customer is likely to cut ties, it can take a more targeted approach to mitigate customer turnover.

In this task, you will use Python, SAS, or R to analyze data for a telecommunications company (see "Customer Data" web link) and create a data mining report in a word processor (e.g., Microsoft Word). You will create visual representations throughout the submission to show each step of your work and to visually represent the findings of your data analysis.

All algorithms and visual representations used need to be captured (either in tables within the word document or with screen shots added into the word document) and should be submitted as part of your document for final submission

A separate Excel (.xls or .xlsx) document of the cleaned data should be submitted along with the written aspects of the data mining report.

SCENARIO

You are an analyst for a telecommunications company that is concerned about the number of customers leaving their landline business for cable competitors. The company needs to know which customers are leaving and attempt to mitigate continued customer loss. You have been asked to analyze customer data to identify why customers are leaving and potential indicators to explain why those customers are leaving so the company can make an informed plan to mitigate further loss.

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. An originality report is provided when you submit your task that 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.

I: Tool Selection

Execute data extraction from the "Customer Data" web link using data mining software (Python, R, or SAS). Provide a screen shot of the code you have written and its successful application with a copy of all the extracted data.

- A. Describe the benefits of using the tool you have chosen (Python, R, or SAS) for extracting data in this scenario.
- B. Define the objectives or goals of the data analysis. Ensure that your objectives or goals are reasonable within the scope of the scenario and are represented in the available data.
- C. Select a descriptive method *and* a nondescriptive method (i.e., predictive, classification, or probabilistic techniques) you will use to analyze the data, and explain how the methods you have selected are appropriate for the objectives or goals you have defined.

II: Data Exploration and Preparation

Clean the data you have extracted and save as .xls or .xlsx format for submission. Be sure to address all necessary formatting, converting, and missing data.

- D. Describe the target variable in the data and indicate the specific type of data the target variable is using, including examples that support your claims.
- E. Describe an independent predictor variable in the data and indicate the specific type of data being described. Use examples from the data set that support your claims.
- F. Propose the goal in manipulation of the data and define your data preparation aims.
- G. Define the statistical identity of the data, including the essential criteria and phenomenon to be predicted.
- H. Explain the steps used to clean the data and how you addressed any anomalies or missing data.

III: Data Analysis

For each of the following steps, be sure to clearly indicate each step within your data sheet with a screen shot and annotations in your final submission. All algorithms used need to be clearly identified in the screen shot and submission.

- I. Identify the distribution of variables using univariate statistics from your cleaned and prepared data. Represent your findings visually as part of your submission.
- J. Identify the distribution of variables using bivariate statistics from your cleaned and prepared data. Represent your findings visually as part of your submission.
- K. Apply an analytic method and an evaluative method. Annotate the data showing both methods and your findings.
- L. Justify the methods you have chosen to analyze your data. Be sure to include details about how the methods you have chosen better represents your findings than other methods.
- M. Justify the methods you have chosen to visually present your data. Be sure to include details about how the presentation methods you chose better represents your findings than other presentation methods.

IV: Data Summary

Summarize the findings of your data evaluation. Provide the final findings dataset, including evaluation measures.

  N. Explain how your data shows that it was discriminating or not and whether the phenomenon you wanted to detect was present in your findings. Provide specific examples from the data to support your claims.

  O. Describe the methods you used for detecting interactions and for selecting the most important predictor variables. Include the specific interactions you detected and the most important predictor variables that you found.

  P. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

File Restrictions

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

File size limit: 200 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

A:TOOL SELECTION: BENEFITS

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not describe benefits of using the tool chosen for extracting data, or a description is not included.	The submission describes benefits of using the tool chosen for extracting data, but does not include specific examples of how this tool is ideal in this scenario over other available tools or the description contains inaccuracies.	The submission describes benefits of using the tool chosen for extracting data, including specific examples of how this tool is ideal in this scenario as opposed to other available tools.

B:TOOL SELECTION: AIMS

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not define any goals or objectives.	The submission defines aims for data analysis, but the goals or objectives are not reasonable or are not based on the scope of the scenario and available represented data.	The submission defines reasonable goals or objectives for data analysis based on the scope of the scenario and available represented data.

C:TOOL SELECTION: PRESCRIBED ANALYSIS

NOT EVIDENT		COMPETENT

The submission does not prescribe a descriptive and a non-descriptive method to be used in the analysis.

APPROACHING COMPETENCE

The submission prescribes the methods to be used in the analysis but does not include an explanation of how the methods are appropriate for the defined aims or the prescribed methods are not appropriate for the defined aims.

The submission prescribes a descriptive and a nondescriptive method to be used in the analysis and explains how the methods selected are appropriate for the defined aims.

D:DATA EXPLORATION AND PREPARATION: TARGET VARIABLE

NOT EVIDENT

The submission does not describe the target variable in the data.

APPROACHING COMPETENCE

The submission describes the target variable in the data but does not indicate the specific type of data the target variable is using with examples from the data set or the description contains inaccuracies.

COMPETENT

The submission describes the target variable in the data, indicating the specific type of data the target variable is using. The submission includes specific examples from the data set to support claims.

E:DATA EXPLORATION AND PREPARATION: INDEPENDENT PREDICTOR

NOT EVIDENT

The submission does not describe the independent predictor variable in the data.

APPROACHING COMPETENCE

The submission describes the independent predictor variable in the data but does not indicate the specific type of data being described. The submission does not include examples from the data set to support claims.

COMPETENT

The submission describes the independent predictor variable in the data and indicates the specific type of data being described. The submission includes examples from the data set to support claims.

F:DATA EXPLORATION AND PREPARATION: GOAL

NOT EVIDENT

The submission does not propose the goal in manipulation of data.

APPROACHING COMPETENCE

The submission proposes the goal in manipulation of data but does not include data preparation aims. Or, the goal or aims proposed are not reasonable or appropriate for the scenario.

COMPETENT

The submission proposes the goal in manipulation of data for this scenario and defines data preparation aims.

G:DATA EXPLORATION AND PREPARATION: STATISTICAL IDENTITY

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not define the data set statistical identity.	The submission defines the data set statistical identity but either does not include the essential criteria and phenomenon to be predicted or contains inaccuracies.	The submission defines the data set statistical identity, including the essential criteria and phenomenon to be predicted.

H:DATA EXPLORATION AND PREPARATION: CLEANING

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not explain the steps used to clean the data.	The submission explains the steps used to clean the data but does not provide both details and evidence of how anomalies or missing data were addressed.	The submission explains the steps used to clean the data and provides details and evidence of how anomalies or missing data were addressed.

I:DATA ANALYSIS: UNIVARIATE STATISTICS

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not identify the distribution of variables using univariate statistics.	The submission identifies the distribution of variables using univariate statistics but does not represent the findings visually. The description contains inaccuracies.	The submission identifies the distribution of variables using univariate statistics from cleaned and prepared data. The findings are represented visually as part of the submission.

J:DATA ANALYSIS: BIVARIATE STATISTICS

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not identify the distribution of variables using bivariate statistics.	The submission identifies the distribution of variables using bivariate statistics but does not represent the findings visually. The description contains inaccuracies.	The submission identifies the distribution of variables using bivariate statistics from cleaned and prepared data. The findings are represented visually as part of the submission.

K:DATA ANALYSIS: METHODS

NOT EVIDENT

The submission does not apply both analytic and evaluative methods.

APPROACHING COMPETENCE

The submission applies both analytic and evaluative methods, but the methods were inaccurately executed. The data is not accurately annotated showing both methods and findings.

COMPETENT

The submission applies both analytic and evaluative methods with accurately annotated methods and findings.

L:DATA ANALYSIS: JUSTIFICATION

NOT EVIDENT

The submission does not justify the methods chosen to analyze the data.

APPROACHING COMPETENCE

The submission justifies the methods chosen to analyze the data but does not include details about how the methods chosen better represent findings for this data set than other methods. The details provided contain inaccuracies.

COMPETENT

The submission justifies the methods chosen to analyze the data, including details about how the method chosen better represents findings for this data set than other methods.

M:DATA ANALYSIS: VISUAL REPRESENTATION

NOT EVIDENT

The submission does not justify the methods chosen to visually present the data.

APPROACHING COMPETENCE

The submission justifies the methods chosen to analyze the data but does not include details about how the presentation method chosen better represents findings for this data set than other methods. The details provided contain inaccuracies.

COMPETENT

The submission justifies the methods chosen to visually present the data and includes details about how the presentation method chosen better represents the findings in this data set than other presentation methods.

N:DATA SUMMARY: PHENOMENON

NOT EVIDENT

The submission does not explain how the data shows discrimination.

APPROACHING COMPETENCE

The submission explains how the data shows it was discriminating or not but either does not explain whether the phenomenon attempted to be detected was present in the findings or does

COMPETENT

The submission explains how the data shows it was discriminating or not and whether the phenomenon attempted to be detected was present in the findings. The submission includes

not provide specific examples from the data to support claims. The explanation contains inaccuracies.

specific examples from the data to support claims.

O:DATA SUMMARY: DETECTION

NOT EVIDENT

The submission does not describe the methods used for detecting interactions and for selecting critical predictor variables.

APPROACHING COMPETENCE

The submission describes the methods used for detecting interactions and for selecting critical predictor variables but does not provide the specific interactions detected and the critical predictor variables that were found. The description contains inaccuracies.

COMPETENT

The submission describes the methods used for detecting interactions and for selecting critical predictor variables. The submission provides the specific interactions detected and the critical predictor variables that were found.

P: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 and/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.

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 student. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar promote accurate interpretation and understanding.

SUPPORTING DOCUMENTS

WA_Fn-UseC_-Telco-Customer-Churn.csv