**C964 Computer Science Capstone**

Ian Rinehart

Western Governors University

C964: Computer Science Capstone

4/8/2024

**Task A:**

**Letter of Transmittal:**

James Roberts

Jim-Bob College

123 Address Road

Salt Lake City, UT 84107

Dear Jim Bob,

I am ready to submit and am happy with the data product I have developed as part of my project to evaluate the impact of changes on male students’ academic performance at Jim-Bob College. This data product aims to provide insight into the effectiveness of recent curriculum changes.

The data product as designed can be used to analyze grade data and predict a student’s gender based on grades. This will assess the impact of the new curriculum on male students. By comparing current students’ grades to historical data, we can determine if these changes have affected male students' performance.

The data product is developed using Python and utilizes a CSV file of our students' information as our data source. It uses machine learning and data visualizations to reach the goals we are aiming for. This will support better decisions about the curriculum in the future.

I am confident that the findings presented in the product will help the ongoing efforts of Jim-Bob College to provide the best education for its students. I look forward to discussing this more in person after the data is collected.

Sincerely,

Ian Rinehart

Lead Developer

Jim-Bob College

**Project Proposal:**

Summary of the Problem:

Women make up almost sixty percent of University students. This itself is not an issue but just 7 years ago men made up seventy-one percent of college students (Hamm, 2021). This means dropouts have gone up significantly. Our school has made significant changes to our curriculum. We wish to see if this has had a change on the male students by comparing their grade data to students' past, and if the answer is correct often then we have not made the best changes.

Description of How the Data Product Benefits the Customer:

The product is crucial for evaluating the effectiveness of the changes we have made to the curriculum. By using the grade data and predicting the students' gender based on grades we can determine if the changes have positively impacted male students who historically lag behind females academically. This will provide good insight to make informed decisions for future improvements to better support male students.

Outline of the Data Product:

The data product is a Python program designed to analyze grades and predict gender based on that. It will be used to gauge the impact of curriculum changes geared towards male students. The program was written in Python 3.12. It can be executed in any Python IDE, I recommend Pycharm as my dependencies for that platform are included. The program utilizes some outside libraries such as pandas, matplotlib, and scikit-learn. The program uses a CSV file as its data source. The program relies on the data from the CSV file, then it performs statistical analysis and machine learning to predict gender based on grades, and finally, the results of this analysis are used to evaluate the impact of the new curriculum. The program generates several graphs to help better understand the data. These show how we value the data in our determinations. The program can be run on any machine with Python installed. The data product provides valuable insights into the impact of curriculum changes on male students' academic performance. It supports informed decision-making processes for educational institutions to improve their curriculum and support systems.

Description of the Data Used to Construct the Data Product:

The data used to create the data product is a CSV file containing information about students' names, gender, race, and their grades in math, reading, and writing. Each row in the CSV file represents a single student, and the columns represent the different attributes.

Objectives and Hypotheses:

The primary objective of this program is to gather data from a student such as their grades to determine their gender. This data can help us make sure the changes we are making at Jim-Bob College are helping our male students catch up. The hypothesis is that the curriculum updates have impacted male students and that the data has changed. By comparing the new data to older data we can determine if the hypothesis is correct and that the gender gap in the classroom has been closed.

Funding Requirements:  
 We need no infrastructure that we do not already have for this simply my staff salary. An exe can be e-mailed to our students and if they wish to take it we can save the data collected for future projects. Our school would need a working e-mail server as well, but Jim-Bob College already has one.

Impact on Stakeholders:

Fewer male dropouts would benefit stakeholders and contribute to a more equitable and inclusive educational environment, ensuring that male students have equal opportunities for academic success. Improved overall student retention rates which are beneficial for the college's reputation and standing. Additionally, reducing male dropouts can positively impact the workforce by increasing the pool of skilled and educated male workers, which can benefit industries and the economy as a whole.

Ethical and Legal Considerations and Precautions with Sensitive Data:

First, we need to obtain consent from these students. This is why the survey is optional. We will be transparent about what we are using this data for with the student's data. We must make sure the data is only available to employees with required access and need to take accountability for following all policies and security making sure individuals who misuse data are held accountable.

My Expertise Relevant to the Solution:

This is my first programming job, but I have been in IT for close to a year, and I grew up knowing a lot about the school system. My mother was a teacher and I grew very familiar with both sides of what educators, leaders, and student go through day to day. I want to help educators make their schools a better experience for all students and using data and machine learning can be a great tool to get us there.

**Task B:**

**Executive Summary:**

Decision Support Problem We are Solving:

The product we are developing addresses the decision support problem of evaluating curriculum changes to Jim-Bob College, and see if it has had a positive impact on boys in particular. The customers are academic institutions in all forms and government policymakers who wish to improve male academic performance. This will accomplish its goal by giving insight into whether the old data still holds up against students' gender today.

The Customers and Why This Product Will Fulfill Their Needs:

As stated the customers are academic institutions and policymakers. They would be interested in improving male student's performance through a new curriculum. As discussed male students are dropping out at alarming rates and their grades lag behind female students (Hamm). This program will provide insight into if their current efforts are working or if the program guesses incorrectly. If the program works perfectly nothing has changed and they need to adjust their education model.

Existing Gaps in Current Data Products:

The existing gaps in the current data products being used lack analysis tools for analysis of specifically male performance. Qualitative data from feedback by students and educators could be of great use. The current product lacks insight into what can help male students improve. This is all room for future growth.

Data Available or Needs to Be Collected:

The data available for this product is a CSV file with students' names, gender, race, and grades. This data provides a good start for analysis, but more data may be needed to do a deep dive analysis as to why these students are falling behind in the classroom. Some of that may also escape the scope of quantitative data. We would however reach a more satisfying and more useful approach by gathering quantitative data.

Methodology Used to Guide and Support Design and Development:

To guide and support design and development I used statistical analysis, machine learning, data training, and data visualization. Using a regression algorithm the program finds the most probable answer to the data. I used this method to take the student's grades and find their gender and charts to visualize the data.

Deliverables Associated With the Design and Development of the Data Product:

The deliverables associated with the project would be a final report detailing findings and recommendations for improvements. Along with that the executable program for collection data along with data visualization. These are designed to provide stakeholders with the data they need to make new decisions. Providing insights into the data to form better-informed decisions.

Plan for Implementation and Anticipated Outcomes:

The implementation plan is to e-mail all students with a school e-mail address (@jimbobcollege.edu). After the students who opened the data product finish their upload we will conduct a comprehensive analysis. We will be developing and testing algorithms and models. After that present findings to stakeholders making recommendations based on the data.

Methods of Validating and Verifying the Product Meets the Needs of the Users and Shareholders:

To make sure we are meeting the needs of our users and shareholders we will conduct peer reviews from our teachers, students, and shareholders. We should also collect feedback from our program making sure it runs without errors. Later down the line, we can simply see if men’s grades have improved.

Programming Environment and Resources Used:

The programming environment I used was the PyCharm Community Edition using the Python programming language. Using Python tools data analysis, machine learning, and data training the program was complete. Along with built-in tools, there were also external libraries such as pandas, matplotlib, and scikit-learn. All of these tools are open-source and require no cost other than my salary and the current e-mail system.

Projected Timeline:

The timeline for this project has milestones to document such as data collection and processing. Afterward, type up all reports and documentation. Then consulting on the data with scholars and peers begins. The project started on April 1, 2024, and the project will likely conclude around April 17, 2024. The timeline depends on several external factors like criticism and complexity of requirements.

**Task D:**

**Documentation:**

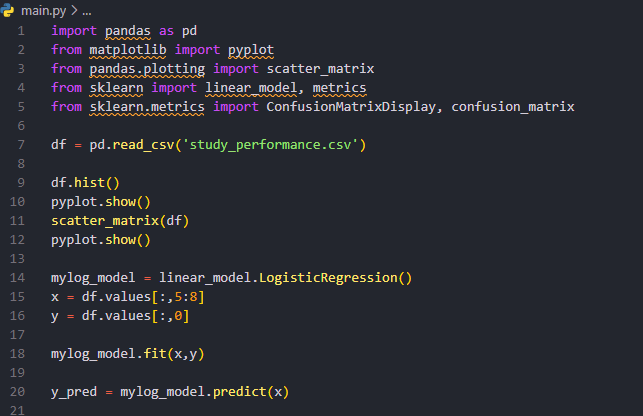
Business Vision/ Requirements:

Jim-Bob College is looking to become a leader in providing inclusive education by improving the academic outcomes of male students. Creating a learning environment for all students, regardless of gender, have equal opportunities for success is our vision. By utilizing the data-driven insight we have access to, we seek to implement effective curriculum changes and support strategies that address the specific needs of male students, ultimately narrowing the achievement gap between male and female college students.

To achieve our vision, we need a data product that can analyze the impact of curriculum changes on male students’ academic performance. This data product will process and analyze student data, like grades and demographics to determine if the student is more likely to be a male or female based on that information. If the data renders incorrect we know the curriculum changes already made recently are having an impact. If not we will take our new data and decide our next action.

Code Used to Perform Analysis:

Using Python which is a great language for machine learning I used some external tools for the data analysis.



The rest of my code was mainly formatting the command line UI. Only twenty lines of code were all my data analysis for the predictive program. In lines 1-7, I import my libraries and CSV data file. In lines 9-12, I show different charts related to the data. Next in lines 14-16 I declare my model and set data values. In the final two lines there I train the data to find the correlation between the X and Y variables and then make a prediction.

Hypothesis Assessment:

Our hypothesis is correct and we do believe curriculum changes are making a positive impact at Jim-Bob College. Through our machine learning algorithm, we have determined that since the program has just over an eighty percent success rate we are making an impact on these students. This success speaks to the importance of data-driven decision-making.

Visualizations and Elements of Effective Storytelling:

The visual elements play a big role in illustrating the impact of the changes on these students’ performance at Jim-Bob College. The visualizations provide a representation of the data before the changes to the curriculum changes. Users can easily view the data as they launch on runtime. Now speaking on reporting our product will tell a compelling story of the impact of curriculum changes on male students’ grades.

Assessment of Product Accuracy:

The data product we developed for evaluating the curriculum changes on male students here at Jim-Bob College has demonstrated a high accuracy level, with an over eighty percent success rate in the gender of the student.

Testing Results:

Product testing is a part of any program life cycle and we take it seriously at Jim-Bob College. First, I and anyone under me will try to iron out any issues ourselves with some critical thinking, but a programmer does not always think about their program as the average user does. Now that bugs were thoroughly tested by our programmers it is time to test with average users. We will deploy a demo for staff members to input their test data and submit bug reports. After bugs are found the team will fix them. Then we would be ready for full deployment.

Source Code:

The user can navigate to my GitHub page for this program at <https://github.com/irinehart/C964-Computer-Science-Capstone.git>. I strongly recommend PyCharm because there already is a setup file for PyCharm so you can just run it easily. There they will see lots of files. Here is a breakdown of what the files are.

* main.py - Source Code
* C964 Computer Science Capstone - This Document
* README.md - Setup Documentation
* study\_performance.CSV - The Data
* Topic Approval - My Topic Approval Form

Quick Start Guide:

(This information will be repeated in the README)

I strongly recommend you use PyCharm to run the program as there are setup files in the project folder, but if you cannot here are some dependencies:

* Python 3.12
* pandas
* matplotlib
* pandas.plotting
* sklearn or sci-kit learn
* sklearn.metrics

After you get the program running a chart will pop up there are three. Simply exit each one out to begin the program. Read the questions carefully and enjoy.

References

Hamm, L. (2021, September 10). Women increasingly outnumber men at U.S. colleges-but

why?. THE FEED. <https://feed.georgetown.edu/access-affordability/women-increasingly-outnumber-men-at-u-s-colleges-but-why/>