D596 Task 1 Writeup

Xavier Loera Flores
ID: 011037676
xloeraf@wgu.edu
D596 The Data Analytics Journey

The Data Analytics Life Cycle

Data Analytics Life Cycle Phases

Business Understanding Phase

Analysts will meet with multiple stakeholders during the business understanding phase to gain a better understanding of the organization's goals for the project.

My expertise in understanding business requirements is very standard for a typical early-stage software engineer where I can understand specific business requirements but may not be able to derive hard requirements and goals from a business mission statement. I plan on improving my expertise in this area by specifically focusing on the needs of nontechnical stakeholders and decision-makers in my conversations at my current job.

Data Acquisition Phase

Data analysts will identify and retrieve data from various sources during the data acquisition phase using either database queries for internal or APIs and web scraping for external data. Once the data is collected, it will be stored for further processing.

I am very capable of querying an internal database or writing scripts to acquire data from an API. I know that my ability to scrape web data or even identify non-readily available data sources is lacking. I plan on improving my expertise in this area by practicing to scrape data from different websites and also making an active effort to identify less obvious data sources.

Data Cleaning Phase

After acquiring the data, analysts need to clean the data to ensure that it is of high quality and free of errors. This may involve fixing improperly formatted data, removing duplicates or missing values, and standardizing data types using a variety of tools and techniques such as Python, R, or SQL.

I have a very low amount of expertise in data cleaning. I have only cleaned invalid symbols from text data in a school project. I definitely need to improve my expertise in this area by studying different data-cleaning techniques for different types of data formats such as images, text, and numerical data.

Data Exploration Phase

Analysts will conduct exploratory data analysis on the cleaned data to gain a better understanding of patterns, trends, and relationships. This phase may involve the use of data visualization tools such as Tableau to help identify key insights and relationships in the data.

I have no experience in exploratory data analysis. I plan on improving my by implementing different data visualization techniques during the development of future projects instead of only creating these visualizations after the fact for a report.

Predictive Modeling Phase

Analysts will use machine learning algorithms to build predictive models that can help stakeholders make informed decisions. This phase may involve the use of tools such as Python and R to help automate the training and use of these models.

I have a very low amount of expertise in predictive modeling as I have only implemented a simple linear regression model in a school project. I plan on improving my expertise in this area by creating more predictive models in future projects and through my own personal study of machine learning algorithms.

Data Mining & Machine Learning Phase

Analysts will use data mining and machine learning techniques to identify patterns and relationships in the resulting data from the predictive models or the data as a whole. Machine learning techniques such as clustering, classification, and regression are implemented to build models that can help stakeholders make informed decisions.

Similar to my expertise in predictive modeling, I have a very low amount of expertise in data mining and machine learning. I need to improve my expertise in this area by building more projects and making a dedicated effort to seek out data mining opportunities in my projects.

Reporting & Visualization Phase

Finally, analysts will present a story of their findings in the form of graphs, reports, and interactive dashboards to key stakeholders. This allows the stakeholders to gain insights from the data to make informed business decisions.

I am very experienced in creating visually rich reports, presentations, and dashboards. Although I am very experienced in this area, I can still improve my abilities by highlighting key insights and relationships in the data with more focus on the business need and mission.

Goal Driven Business Requirements Identification

Business missions and goals help analysts identify business requirements by providing a clear understanding of the organization's objectives. It not only serves as a primer for understanding the needs of the stakeholders but it also serves as a north start for the data analytics cycle. An analyst can ensure their work aligns with the organization's goals and objectives so long as their decisions serve the mission statement.

Application of a Data Analytics Tool in the Data Analytics Life Cycle

At my current company, I have the opportunity to potentially improve student success outcomes using student data. In the predictive modeling phase of the data analytics life cycle, I can apply logistic regression to predict student success based on features like assignment scores, quiz performance, and class attendance potentially identifying and helping at-risk students early on.

Risks

There are risks involved in using predictive models to forecast student outcomes. Some of these risks include:
- There may be data quality issues if student progress data is incomplete or inaccurate since it is manually tracked by the instructional team making predictions less reliable.
- There is a privacy risk if the model is not properly secured and student data is exposed to unauthorized users thus violating potential FERPA regulations.
- The model could unintentionally be biased against certain cohorts or subgroups of students. We will need to ensure that the model accounts for varying learning paces and doesn't penalize students who learn differently.

Organizational Problem

The proposed predictive model aims to solve the problem of identifying at-risk students early on allowing the instructional team to intervene and fix the course of the student's success outcome. By using predictive analytics based on student performance data, the model addresses my company's need to improve student success outcomes ensuring that students and educators are successful. Increasing student success outcomes not only helps reduce the number of students dropping out and refunding the course but also helps the company's reputation with students and educators.

Data Analytics Decision Making Process

The decision-making process for implementing a predicted model began with gaining a business understanding of the problem. Once the problem was understood, exploring the data that was available

helped identify potential features that could be used to predict student success. Given our available data and business needs, logistic regression was chosen as a predictive model to forecast student success.

Justification

My company needs the predictive model because it will help identify at-risk students early on and provide the instructional team with the necessary information to intervene and help the student succeed. Logistic regression works well for our situation since we have datasets with features such as assignment scores, quiz performance, and attendance. The classification results would also be fast and easy to interpret and act upon.

Results

The results of using the predicted model will be a categorized list of students who are at risk and need intervention. These models can be run multiple times at different stages of the course to ensure that the instructional team is always up to date with the student's progress. The results can also be used to evaluate the effectiveness of the instructional team or previous interventions.

Potential Ethical Problems

There are ethical concerns with our predictive model solution including:
- **Data Privacy**: The model risks exposing sensitive student information which may violate FERPA privacy laws.
- **Bias**: The model may unintentionally discriminate against certain student groups which may result in unfair interventions and unaddressed students who were falsely marked as doing well.
- **Psychological Consequences**: The model might reduce educators' judgment if they make decisions based solely on predictions rather than individual student needs. There is also concern about how the students will react to being labeled as at-risk.