Capstone Assessment: Data Analytics

*Masyn Grisel 2/18/2025*

1. Competencies, Skills, and Knowledge (What I Know)

Course-Based Learning

Course: CIDM 6308 – Seminar in Data Analytics

* Key Topics Covered:
  + Logistic regression modeling
  + Data cleaning and preprocessing
  + SQL queries and joins for data retrieval
  + Predictive modeling for decision-making using RapidMiner and R
  + Data Analytics using Python
* Examples of Work:  
  Project 1: Data Cleaning and Predictive Modeling with RapidMiner
  + Objective: Process and clean raw data to prepare it for analysis.
  + Tools Used: RapidMiner, Excel
  + Process:
    - Imported a dataset with missing values and inconsistencies.
    - Applied data preprocessing techniques (handling missing values, normalizing data, removing duplicates).
    - Used logistic regression in RapidMiner to predict outcomes based on cleaned data.
  + Outcome: Improved data quality and developed a predictive model for decision-making.

Project 2: SQL-Based Data Analysis and Decision Support

* + Objective: Extract meaningful insights from a structured database using SQL.
  + Tools Used: SQL (MySQL)
  + Process:
    - Used SQL joins (INNER JOIN, LEFT JOIN, etc.) to combine tables and retrieve necessary data.
    - Wrote queries with WHERE, GROUP BY, and ORDER BY to analyze specific trends.
    - Applied subqueries and aggregate functions to summarize and filter data.
  + Outcome: Successfully identified key business trends and provided actionable insights based on SQL queries.

Technical Skills & Tools Proficiency

* Excel: Pivot tables, slicers, dashboards, statistical analysis
* SQL: Querying and managing datasets
* Python/R: Pandas, Numpy, matplotlib, ggplot2
* Machine Learning: Logistic regression, decision tree models, k-means clustering, neural networks, linear regression
* Data Visualization: Tableau, Python

2. Weaknesses (Where I Am Least Confident)

* Advanced machine learning techniques
* Real world predictive modeling
* Memory of libraries and tools(Python libraries, syntax, identifying the correct type of model)

3. Gaps in Knowledge (What I Wish I Knew or Might Be Missing)

* Using cloud-based analytics tools
* Using big data tools
* Deeper understanding of the ETL process

4. Sources of Knowledge & Learning Materials

* DataCamp: Courses in SQL, Python, Tableau, and Excel
* Lecture Videos: Step by step instructions provided by professor Chen on how to use certain tools(RapidMiner and Excel)
* YouTube: Any time I had a minor issue that was not worth emailing the professor, YouTube often had a good explanation.

5. Summary Statement & Preparedness

I have developed a solid foundation in data analytics, particularly in areas such as data cleaning, basic machine learning models, data visualization, and SQL querying. However, I recognize that there are key areas in which I need to further build my knowledge and skills to be fully prepared for real-world data analytics challenges.

Overall Strengths

I am confident in my ability to handle data preparation tasks such as data cleaning. My understanding of logistic regression and decision trees has allowed me to implement basic predictive models, and I am skilled in using SQL for extracting meaningful insights from structured datasets. I am also skilled in cleaning data in Excel and creating pivot tables or using V-Lookup. Additionally, I have solid experience with Python libraries like Pandas, NumPy, and matplotlib, which I’ve used for tasks such as data manipulation and visualization.

Key Areas for Growth

There are areas where I feel less confident and need further development:

1. Advanced Machine Learning Techniques – I have not yet fully explored more complex algorithms such as support vector machines, ensemble learning, or deep learning. Gaining proficiency in these techniques would enable me to tackle a wider range of data problems and improve model accuracy.
2. Real-World Predictive Modeling – While I have worked on predictive modeling in coursework, I have limited experience with real-world data, which often presents challenges like missing values, noisy inputs, and domain-specific constraints. More exposure to these scenarios would help me understand how to develop models that are both accurate and practical for business use.
3. Memory of Libraries and Tools – As I continue to learn and use Python and R, I need to strengthen my familiarity with the syntax of various libraries and become quicker at identifying the right model for specific datasets. Improving this area will increase my efficiency in applying machine learning techniques and improve my confidence in tackling more complex projects.

How I Can Apply This Knowledge

The skills I have developed in data analysis, machine learning, and visualization will be valuable for my Capstone project and future work. By focusing on advanced machine learning techniques, real-world predictive modeling, and improving my memory of tools and libraries, I will be well-equipped to create more robust data models and present actionable insights. I am confident that addressing these gaps will not only strengthen my technical capabilities but also enhance my ability to make data-driven decisions in professional settings.