Deliverable 1.2 – Alternative Learning Pathways per Learning Unit

Introduction

Deliverable 1.2 – Alternative Learning Pathways per Learning Unit, presents a systematically designed framework of competencies and corresponding learning trajectories for the IM0503 Data Analytics course. It serves as a foundational component of the EduVerse project as an enabler of the adaptive learning pedagogical approach.

Specifically, this deliverable outlines a comprehensive, hierarchical structure of learning competencies, categorized into Foundational Data Analytics Understanding, Data Exploration and Preparation, Core Machine Learning Applications, and Text Analytics. Each competency is defined, detailing both conceptual knowledge and practical tool proficiency, where applicable. Crucially, the report then maps these competencies onto three distinct alternative learning pathways: Business Background, Technical Background, and the Accelerated Learners. For each pathway, we address the specific learning needs (See Deliverable 1.1 – List of Learning Needs) of their corresponding target audience, providing justifications based on their backgrounds and pedagogical requirements.

EduVerse aims to enhance education through an adaptive pedagogical model, utilizing Artificial Intelligence (AI) and Augmented Reality (AR) to personalize the learning experience. This framework of alternative learning pathways contributes to Work Package 2 (WP2). These defined pathways can be used to dynamically profile individual students, assess their prior knowledge and learning goals, and then prescribe the most suitable learning trajectory within the IM0503 course. This ensures that each student receives content and activities optimally aligned with their needs, avoiding redundancy for advanced learners and providing necessary scaffolding for those with different backgrounds.

Beyond its direct application within EduVerse, Deliverable 1.2 is usable for other education professionals seeking to innovate their teaching practices. It provides a concrete example and an applicable model for designing adaptive curricula, enhancing personalized learning initiatives, informing instructional design, and facilitating technology integration into education.

IM0503 Data Analytics Learning Objectives

- Understand the essential concepts in machine learning, such as a dataset, algorithm, model, and accuracy,
- Understand the key concepts of data analytics and their role for organizations in decision-making and innovation,
- Apply the CRISP-DM methodology, which is an industry standard to solve data mining tasks,
- Understand the steps and stakeholders in the knowledge discovery and data mining processes,
- Analyze the needs of an organization and transform them into a data analytics task,
- Apply the following analytical techniques: multiple regression, clustering (e.g., K-means), classification (e.g., Naïve Bayes, KNN, decision trees), dimensionality reduction (e.g., principal component analysis), association rule mining, and text analytics,
- Evaluate the outcomes of data analytics processes and describe them to various stakeholders,
- Use a process-based data analytics tool such as RapidMiner,
- Understand the difference between predictive, prescriptive, and descriptive analytics and be able to decide which one to use in a given situation.

IM0503 Data Analytics Course Competency Framework

This competency framework is structured hierarchically to reflect prerequisite knowledge. Education professionals can use this structure to identify their students' current knowledge gaps and focus on the most relevant learning units, enabling a personalized adaptive learning experience. "Tool Proficiency" competencies are specific applications using RapidMiner, which is the primary tool emphasized in the workbook.

The competencies are categorized under four groups: I. Foundational Data Analytics Understanding, II. Data Exploration and Preparation, III. Core Machine Learning Applications, and IV. Text Analytics. The competencies and their explanations are elaborated in this deliverable, along with their dependencies.

I. Foundational Data Analytics Understanding

(These competencies form the foundation for all subsequent learning. A strong grasp here is crucial for anyone pursuing data analytics. See Figure 1)

C1.1: Understand Data Analytic Mindset

- **Explanation:** Grasping the overarching philosophy of data analytics, including the "Data to Wisdom" hierarchy, and recognizing the various roles within a data analytics team.
- **Dependencies:** None. This is a foundational, conceptual starting point.

C1.2: Differentiate Data Types and Their Operations

- **Explanation:** Distinguishing between categorical (nominal, ordinal) and numerical (interval, ratio) data, and understanding which statistical and mathematical operations are meaningful for each type.
- **Dependencies:** C1.1 (Requires a basic understanding of what "data" is to categorize it).

C1.3: Identify Data Sources

- **Explanation:** Recognizing various internal and external organizational systems and public datasets as potential sources for data analytics.
- **Dependencies:** C1.1 (Building on the fundamental concept of "data" and its relevance).

C1.4: Comprehend Machine Learning Paradigms

- **Explanation:** Understanding the fundamental differences between supervised, unsupervised, semi-supervised, self-supervised, and reinforcement learning, along with their core strengths and weaknesses.
- **Dependencies:** C1.1 (Requires a basic understanding of "algorithms" and "models" as introduced in the mindset).

C1.5: Apply CRISP-DM Process

- **Explanation:** Systematically following the Cross-Industry Standard Process for Data Mining (CRISP-DM) steps: business understanding, data understanding, data preparation, modelling, and evaluation.
- **Dependencies:** C1.1 (CRISP-DM provides a structured approach to the overall data analytic mindset).

II. Data Exploration and Preparation

(These competencies enable students to handle raw data, identify quality issues, and transform data into a suitable format for analysis and modeling.)

C2.1: Perform Basic Data Visualization

- **Explanation:** Creating and interpreting common charts (line charts, scatter plots, histograms, pie charts, box plots, bubble charts) to explore data patterns and communicate insights.
- **Dependencies:** C1.2 (Choosing appropriate visualizations heavily depends on understanding data types).
- **C2.1T: Use RapidMiner for Basic Data Visualization**
    - **Explanation:** Practical application of RapidMiner to generate the aforementioned basic charts.
    - **Dependencies:** C2.1 (Conceptual understanding of visualization types) and assumed basic RapidMiner interface familiarity (can be learned concurrently).

C2.2: Identify Data Quality Issues

- **Explanation:** Recognizing common problems in datasets such as missing values, outliers, inconsistent scales/units, and noise.
- **Dependencies:** C1.2 (An understanding of data characteristics is essential to spot anomalies or inconsistencies). C2.1 (Visualization often helps in early identification of issues).

C2.3: Handle Missing Values

- **Explanation:** Implementing strategies to effectively remove or impute missing values to maintain data integrity for modeling.
- **Dependencies:** C2.2 (One must first identify missing values before handling them).
- **C2.3T: Use RapidMiner for Missing Value Handling**
  - **Explanation:** Practical application of RapidMiner operators to remove or replace missing values.
  - **Dependencies:** C2.3 (Conceptual understanding).

C2.4: Detect and Manage Outliers

- **Explanation:** Employing visual inspection (e.g., box plots) and programmatic methods (e.g., z-score, IQR) to detect outliers and deciding on appropriate data treatment.
- **Dependencies:** C2.2 (Identification of outliers as a quality issue), C2.1 (Box plots are a visualization tool).
- **C2.4T: Use RapidMiner for Outlier Detection**
  - **Explanation:** Practical application of RapidMiner operators like "Detect Outliers (Univariate)".
  - **Dependencies:** C2.4 (Conceptual understanding).

C2.5: Apply Feature Scaling

- **Explanation:** Understanding the purpose and applying techniques like normalization and standardization to bring attributes to a similar scale, and knowing when each is suitable for different machine learning algorithms.
- **Dependencies:** C2.2 (Different scales are a data quality issue), C1.4 (Understanding how ML algorithms are affected by scale differences).
- **C2.5T: Use RapidMiner for Feature Scaling**
  - **Explanation:** Practical application of RapidMiner's "Normalize" operator with various methods.
  - **Dependencies:** C2.5 (Conceptual understanding).

C2.6: Perform Discretization

- **Explanation:** Transforming continuous numerical data into discrete, categorical bins or intervals.
- **Dependencies:** C1.2 (Understanding the distinction between continuous and categorical data).
- **C2.6T: Use RapidMiner for Discretization**
  - **Explanation:** Practical application of RapidMiner's "Discretize" operator.
  - **Dependencies:** C2.6 (Conceptual understanding).

C2.7: Implement Dimensionality Reduction

- **Explanation:** Applying techniques like feature selection (e.g., removing highly correlated features) and Principal Component Analysis (PCA) to reduce the number of attributes while retaining most relevant information.
- **Dependencies:** C1.2 (Understanding attributes/features), C2.2 (Recognizing redundancy as a data issue).
- **C2.7T: Use RapidMiner for Dimensionality Reduction**
  - **Explanation:** Practical application of RapidMiner's "Remove Correlated Attributes" and "PCA" operators.
  - **Dependencies:** C2.7 (Conceptual understanding).

III. Core Machine Learning Applications

(This section covers the central machine learning paradigms, their algorithms, and how to evaluate their performance.)

C3.1: Differentiate Analytics Types (Descriptive, Predictive, Prescriptive)

- **Explanation:** Understanding the distinct goals, value, and implementation difficulty of descriptive, predictive, and prescriptive analytics.
- **Dependencies:** C1.1 (Builds on the general purpose of data analytics).

C3.2: Understand Prediction Model Training and Deployment

- **Explanation:** Grasping the concepts of splitting data into training, validation, and holdout sets, and the overall process from model training to deployment.
- **Dependencies:** C1.4 (Requires understanding of supervised learning).

C3.3: Evaluate Model Performance and Generalizability (General)

- **Explanation:** Understanding general concepts of model evaluation, the importance of generalizability, and identifying issues like underfitting and overfitting.
- **Dependencies:** C3.2 (Evaluation is performed on trained models), C1.4 (General understanding of ML model behavior).

C3.4: Apply Supervised Classification Algorithms

- **Explanation:** Implementing and understanding supervised classification algorithms, specifically Decision Trees (using Information Gain) and K-Nearest Neighbors (KNN).
- **Dependencies:** C1.4 (Understands supervised learning), C2.5 (Feature scaling is often crucial for KNN), C2.7 (Dimensionality reduction can improve classification).
- **C3.4T: Use RapidMiner for Classification**
  - **Explanation:** Practical application of RapidMiner operators for Decision Trees and k-NN.
  - **Dependencies:** C3.4 (Conceptual understanding).

C3.5: Evaluate Classification Performance

- **Explanation:** Using metrics such as Confusion Matrix (TP, FP, FN, TN), Precision, Recall, F-1 Score, and Expected Value (cost/benefit reasoning) to assess the performance of classification models, including multiclass scenarios.
- **Dependencies:** C3.3 (General evaluation concepts), C3.4 (Requires classification model output).

C3.6: Apply Supervised Regression Algorithms

- **Explanation:** Implementing and understanding supervised regression algorithms, specifically Linear Regression, and its underlying formula.
- **Dependencies:** C1.4 (Understands supervised learning), C2.5 (Feature scaling is often critical for regression models).
- **C3.6T: Use RapidMiner for Regression**
  - **Explanation:** Practical application of RapidMiner's "Linear Regression" operator.
  - **Dependencies:** C3.6 (Conceptual understanding).

C3.7: Evaluate Regression Performance

- **Explanation:** Using objective functions like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Median Absolute Deviation (MAD) to evaluate regression model performance.
- **Dependencies:** C3.3 (General evaluation concepts), C3.6 (Requires regression model output).

C3.8: Apply Unsupervised Clustering Algorithms

- **Explanation:** Implementing and understanding unsupervised clustering algorithms, including hierarchical, objective function-based (e.g., K-means), density-based (e.g., DBSCAN), and grid-based clustering.
- **Dependencies:** C1.4 (Understands unsupervised learning), C2.5 (Feature scaling often required for distance-based clustering).
- **C3.8T: Use RapidMiner for Clustering**
  - **Explanation:** Practical application of RapidMiner's "Clustering (k-Means)" and other clustering operators.
  - **Dependencies:** C3.8 (Conceptual understanding).

C3.9: Evaluate Clustering Performance

- **Explanation:** Using metrics like Silhouette Analysis, Dunn Index, and Davies-Bouldin Index to evaluate the quality and separation of clusters.
- **Dependencies:** C3.3 (General evaluation concepts), C3.8 (Requires clustering model output).

C3.10: Apply Association Rule Mining

- **Explanation:** Implementing and understanding algorithms like Apriori and FP-Growth to discover patterns of co-occurrence in transactional data, utilizing concepts of support, confidence, and lift.
- **Dependencies:** C1.4 (Understands unsupervised learning).

IV. Text Analytics
(This specialized section focuses on applying data analytics techniques to unstructured text data, from preprocessing to advanced modeling.)

C4.1: Understand Text as Data and Its Challenges

- **Explanation:** Recognizing the unique unstructured nature of text data, its diverse sources, and fundamental concepts like corpus, document, and tokens, along with associated challenges.
- **Dependencies:** C1.1 (Extends general data understanding to a specific data type).

C4.2: Perform Text Preprocessing

- **Explanation:** Applying essential techniques such as tokenization, text normalization (e.g., lowercase conversion, punctuation removal), stemming, lemmatization, and stopword removal to prepare text for analysis.
- **Dependencies:** C4.1 (Requires an understanding of raw text data and its inherent "noise").

C4.3: Implement Text Vectorization

- **Explanation:** Transforming preprocessed text into numerical vector representations using methods like one-hot encoding, Bag-of-Words (binary, term count, term frequency), TF-IDF, and N-grams.
- **Dependencies:** C4.2 (Vectorization typically requires preprocessed text), C1.2 (Understanding how to convert qualitative data into numerical forms).

C4.4: Measure Text Similarity

- **Explanation:** Calculating similarity and distance between text documents using metrics such as Euclidean, Manhattan, Minkowski distances, and Cosine Similarity.
- **Dependencies:** C4.3 (Text must be in a vectorized numerical format to calculate similarity).

C4.5: Apply Text Classification

- **Explanation:** Utilizing machine learning algorithms (e.g., KNN, Logistic Regression, SVM, Naïve Bayes) for tasks like categorizing news articles, prioritizing support tickets, or sentiment analysis.
- **Dependencies:** C3.4 (Applies general classification algorithms), C4.3 (Requires vectorized text data as input for ML models).

C4.6: Apply Text Clustering and Topic Modelling

- **Explanation:** Grouping similar text documents using clustering algorithms and understanding how Topic Modelling can identify underlying themes in a corpus.
- **Dependencies:** C3.8 (Applies general clustering algorithms), C4.3 (Requires vectorized text data as input).

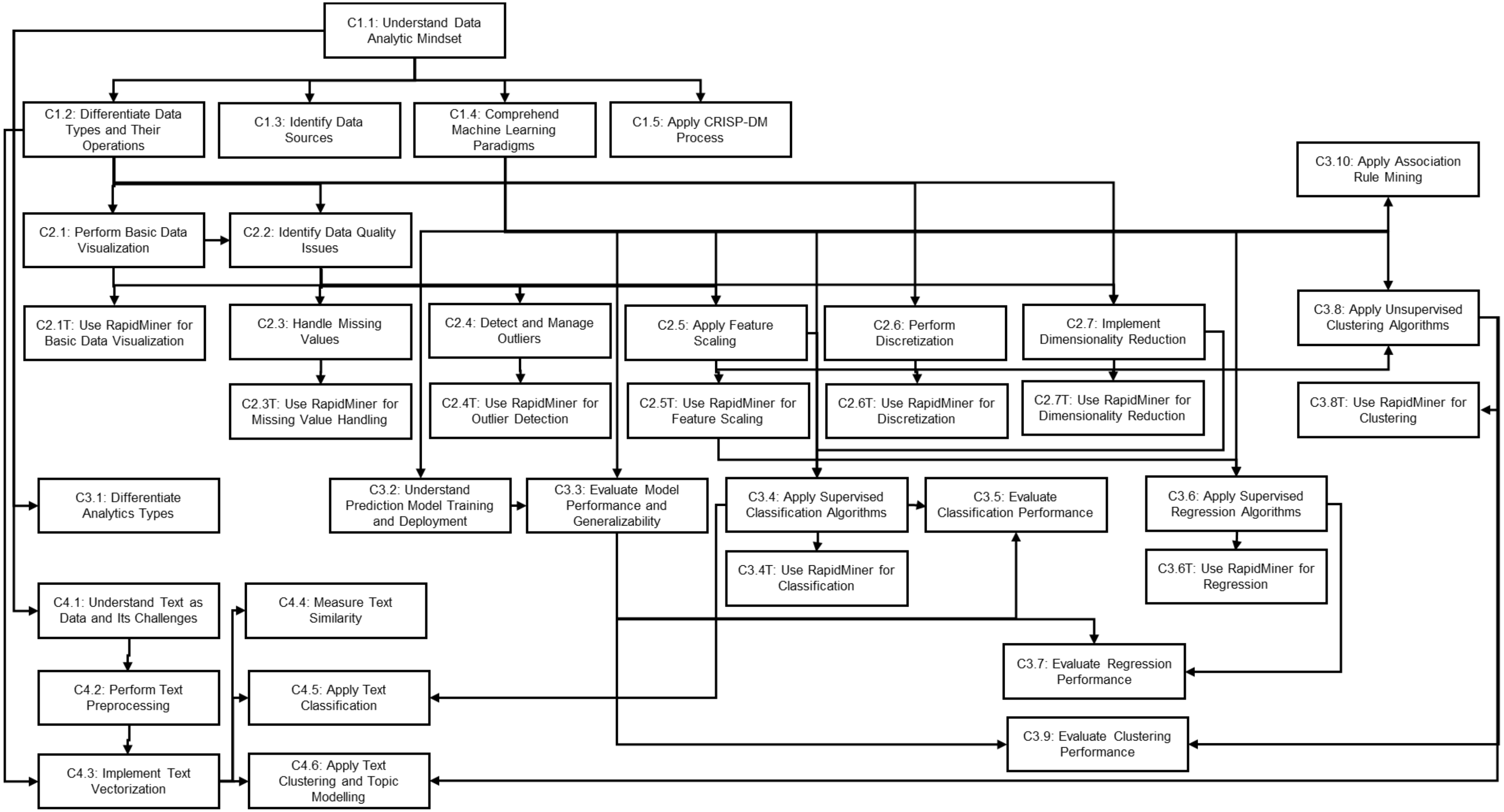

Figure 1: Hierarchical competence structure of IM0503 Data Analytics

Adaptive Learning Pathways

These pathways address different student profiles, allowing them to focus on areas most relevant to their backgrounds and career aspirations.

Pathway 1: Business Background (Focus on Insights & Decision Making)

See Figure 2

- **Target Audience:** Students from business administration, management, marketing, or non-technical backgrounds who aim to understand how data analytics can drive strategic decisions. They need to interpret results, identify opportunities, and communicate with technical teams, rather than focusing on deep implementation.
- **Starting Point:** C1.1 (Understand Data Analytic Mindset).
- **Pathway Flow:**
  - **Foundational (I):** Complete all (C1.1-C1.5). *Emphasis on C1.5 (CRISP-DM)* for project structuring and *C1.4 (ML Paradigms)* to understand capabilities.
  - **Data Exploration & Preparation (II):** Complete C2.1 (Basic Data Visualization) and C2.2 (Identify Data Quality Issues). *Lighter emphasis* on the detailed technical steps of C2.3-C2.7 (Missing Values, Outliers, Scaling, Discretization, Dimensionality Reduction) and their RapidMiner applications (C2.xT). Focus is on *understanding why these steps are necessary for data quality and model reliability*, rather than hands-on implementation.
  - **Core Machine Learning Applications (III):** Complete C3.1 (Differentiate Analytics Types), C3.2 (Understand Prediction Model Training), and C3.3 (Evaluate General Performance). *Strong emphasis* on C3.5 (Evaluate Classification Performance, especially Expected Value/Cost-Benefit) and C3.7 (Evaluate Regression Performance) for decision-making. Review C3.4, C3.6, C3.8, C3.10 to grasp *what each algorithm does* and *what problems it solves*, rather than its internal mechanics. The RapidMiner tool competencies (C3.xT) are for conceptual demonstration and result interpretation.
  - **Text Analytics (IV):** Complete C4.1 (Understand Text as Data). *Strong emphasis* on C4.5 (Apply Text Classification for sentiment, topic) and C4.6 (Text Clustering/Topic Modelling for insights). *Lighter emphasis* on detailed C4.2 (Preprocessing), C4.3 (Vectorization), and C4.4 (Similarity calculation technicalities).

Pathway 2: Technical Background (Focus on Algorithms & Tooling)

See Figure 3

- **Target Audience:** Students from computer science, information science, engineering, or related fields with strong programming/mathematical foundations. They aim to build, implement, and optimize data analytics solutions.
- **Starting Point:** C1.1 (Quick review if prior knowledge exists).
- **Pathway Flow:**
  - **Foundational (I):** Quick review of C1.1, C1.3, C1.5. *Deep dive into C1.2 (Data Types)* and *C1.4 (ML Paradigms)* for foundational technical understanding.
  - **Data Exploration & Preparation (II):** Complete all (C2.1-C2.7). *Strong emphasis* on the technical details of C2.xT (RapidMiner implementation), including parameter tuning and understanding the *mathematical rationale* behind C2.4 (Outlier Detection), C2.5 (Feature Scaling), and C2.7 (Dimensionality Reduction).

- o **Core Machine Learning Applications (III):** Complete all (C3.1-C3.10). *Deep dive into the internal workings and mathematical foundations* of all algorithms (C3.4, C3.6, C3.8, C3.10) and their specific RapidMiner implementations (C3.xT). *Strong emphasis* on the formulas and calculations behind performance metrics (C3.5, C3.7, C3.9).
- o **Text Analytics (IV):** Complete all (C4.1-C4.6). *Deep dive into the technicalities* of C4.2 (Text Preprocessing methods), C4.3 (Vectorization algorithms like TF-IDF, N-grams), and the *mathematics of C4.4 (Text Similarity)*. Focus on implementing and optimizing text analytics pipelines.

Pathway 3: Accelerated Learners (For those with prior foundational knowledge)

See Figure 4

- **Target Audience:** Students who already possess a good understanding of foundational ML concepts and basic data preparation (e.g., from an undergraduate course, extensive MOOCs, or professional experience). They want to quickly get to the advanced applications and specific RapidMiner implementations.
- **Starting Point:** Self-assessment. If proficient in I & II, they can directly start with **C3.1 (Differentiate Analytics Types)** or even **C3.4 (Apply Classification Algorithms)**.
- **Pathway Flow:**
  - o **Foundational (I):** *Self-assessment and quick review* of C1.1-C1.5. Only spend time on areas identified as weak.
  - o **Data Exploration & Preparation (II):** *Self-assessment and focused review*. Can *skip C2.1 (Basic Data Visualization)* if proficient. Can *review C2.2-C2.7 and their T versions* to align with RapidMiner's specific operators and ensure conceptual understanding translates to the tool.
  - o **Core Machine Learning Applications (III):** Complete all (C3.1-C3.10). Focus on the *specific algorithms covered in this course* (Decision Trees, KNN, Linear Regression, Apriori, FP-Growth, Clustering types) and their *RapidMiner implementation details (C3.xT)*. Deepen understanding of how to interpret and tune parameters, and the nuances of the performance evaluation metrics (C3.5, C3.7, C3.9).
  - o **Text Analytics (IV):** Complete all (C4.1-C4.6). Concentrate on how the general ML concepts apply to text data and master the *specific text-processing and vectorization techniques* detailed in the course (C4.2, C4.3, C4.4). Explore the application of classification and clustering to text.

IM0503 Learning Units

Following table presents a mapping between the learning units in which IM0503 Data Analytics course is structured and the competencies covered by those learning units.

Table 1: Learning Units and corresponding competencies

Learning Unit (LU)	**Competencies**
LU1: Data Analytic Thinking	C1.1, C1.2, C1.3, C1.4, C1.5
LU 2: Data Visualization	C2.1, C2.1T
LU 3: Data Quality and Data Preparation	C2.2, C2.3, C2.3T, C2.4, C2.4T, C2.5, C2.5T, C2.6, C2.6T, C2.7, C2.7T
LU 4: Fundamentals of Machine Learning	C3.1, C3.2, C3.3
LU 5: Classification	C3.4, C3.4T, C3.5
LU 6: Regression	C3.6, C3.6T, C3.7
LU 7: Clustering	C3.8, C3.8T, C3.9
LU 8: Association Rule Mining	C3.10
LU 9: Text Analytics	C4.1, C4.2, C4.3, C4.4, C4.5, C4.6

Mapping between Learning Needs and Competencies

Following table presents a mapping between the learning needs identified and elaborated in Deliverable 1.1 List of Learning Needs and the competencies.

Table 2: Mapping between Learning Needs and Competencies

Learner Category	**Learning Needs**	**Competencies**
Business Background	1. Conceptual Understanding of ML, AI, and Analytics Types	C1.1, C1.4, C3.1
	2. CRISP-DM as a Project Management Framework	C1.5
	3. Data Quality Awareness & Business Impact	C2.2
	4. Interpretation of Data Visualizations	C2.1
	5. Evaluation Metrics for Business Objectives & Trade-offs	C3.3, C3.5, C3.7, C3.9
	6. Understanding Algorithm Applications & Limitations	C3.4, C3.6, C3.8, C3.10, C4.5, C4.6

Technical Background	1. Deep Mathematical & Statistical Foundations	C1.2, C1.4, C2.4, C2.5, C3.5, C3.7, C3.9, C4.4
	2. Proficiency in Data Preparation Techniques	All C2.x, especially C2.xT
	3. Algorithm Mechanics and Implementation	C3.4, C3.6, C3.8, C3.10, C4.5, C4.6, including C3.xT
	4. Rigorous Model Evaluation & Validation	C3.3, C3.5, C3.7, C3.9
	5. Text Processing and Vectorization Mechanics	C4.2, C4.3, C4.4
	6. Troubleshooting & Debugging Analytical Workflows	C1.1, C3.3
Accelerated Learner	1. Efficient Gap Identification & Targeted Learning	
	2. RapidMiner Tool Proficiency for Specific Implementations	All C.xT versions
	3. Nuances of Algorithm Selection, Tuning, and Performance	C3.4, C3.6, C3.8, C3.10
	4. Advanced Evaluation Strategies & Business Alignment	C3.5, C3.7, C3.9
	5. Integration of Specialized Topics	C4.x
	6. Learning Efficient Workflows and Best Practices	All C.xT versions

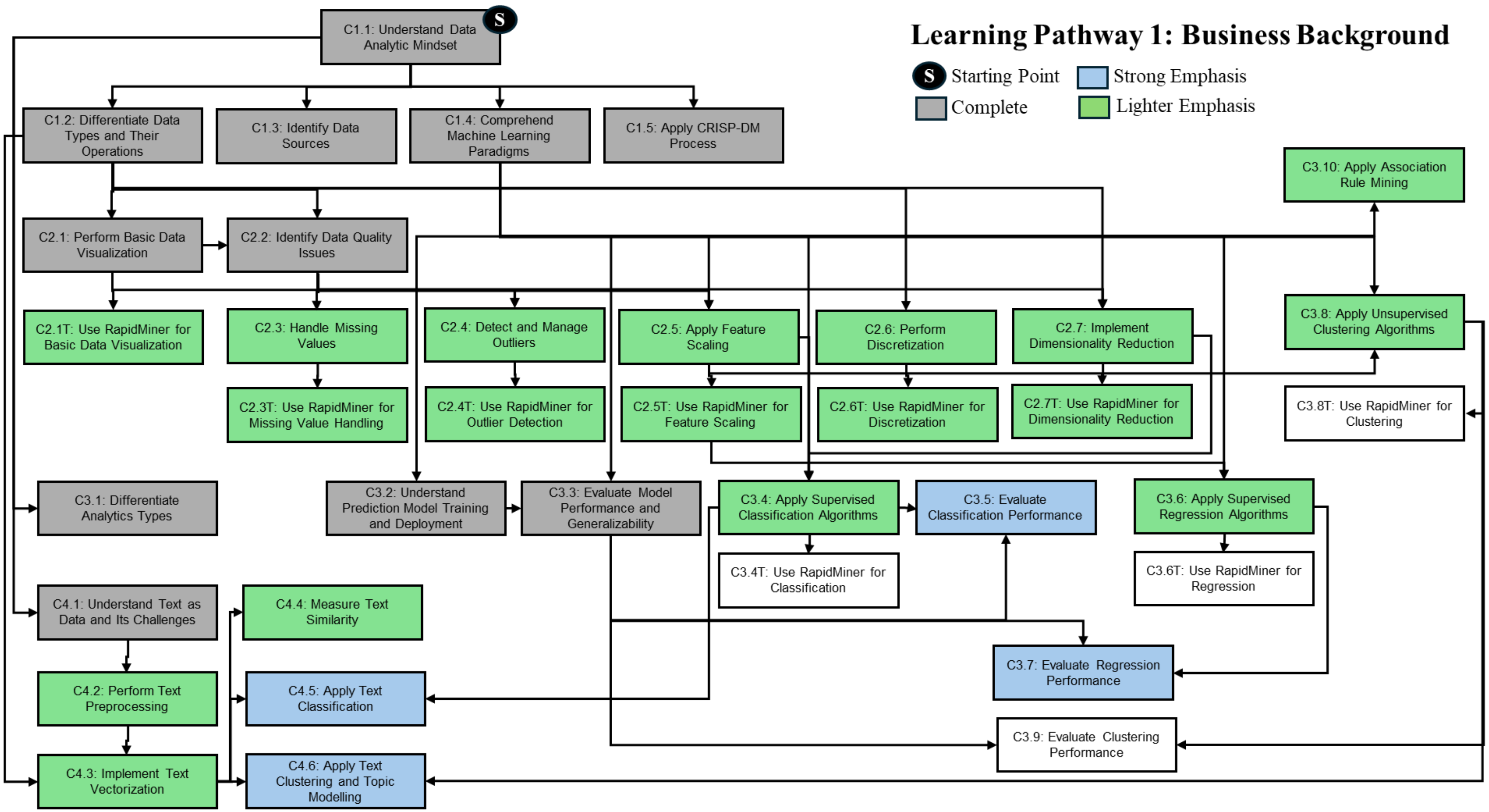

Figure 2: Visual representation of Learning Pathway 1 - Business Background

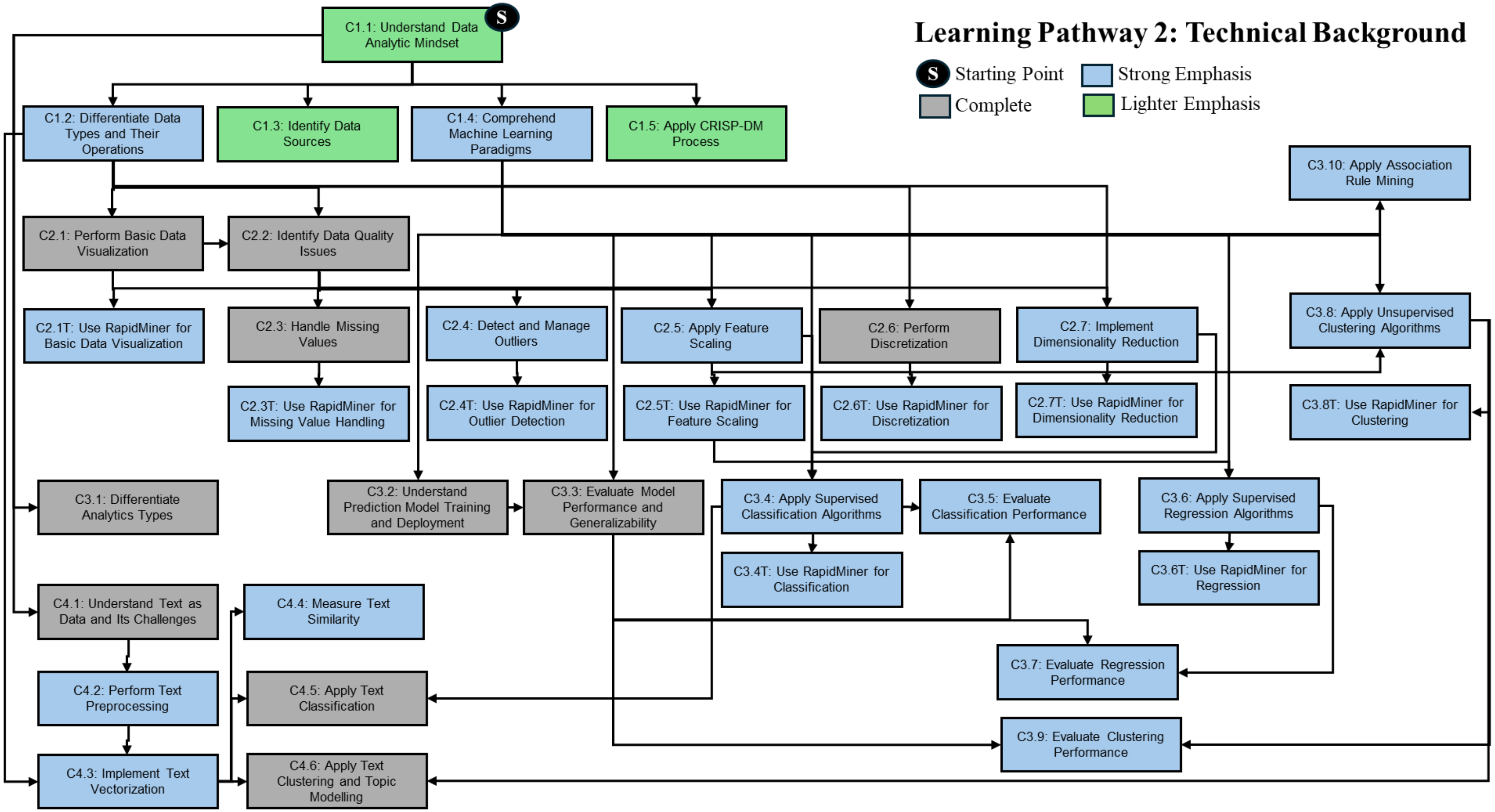

Figure 3: Visual representation of Learning Pathway 2 - Technical Background

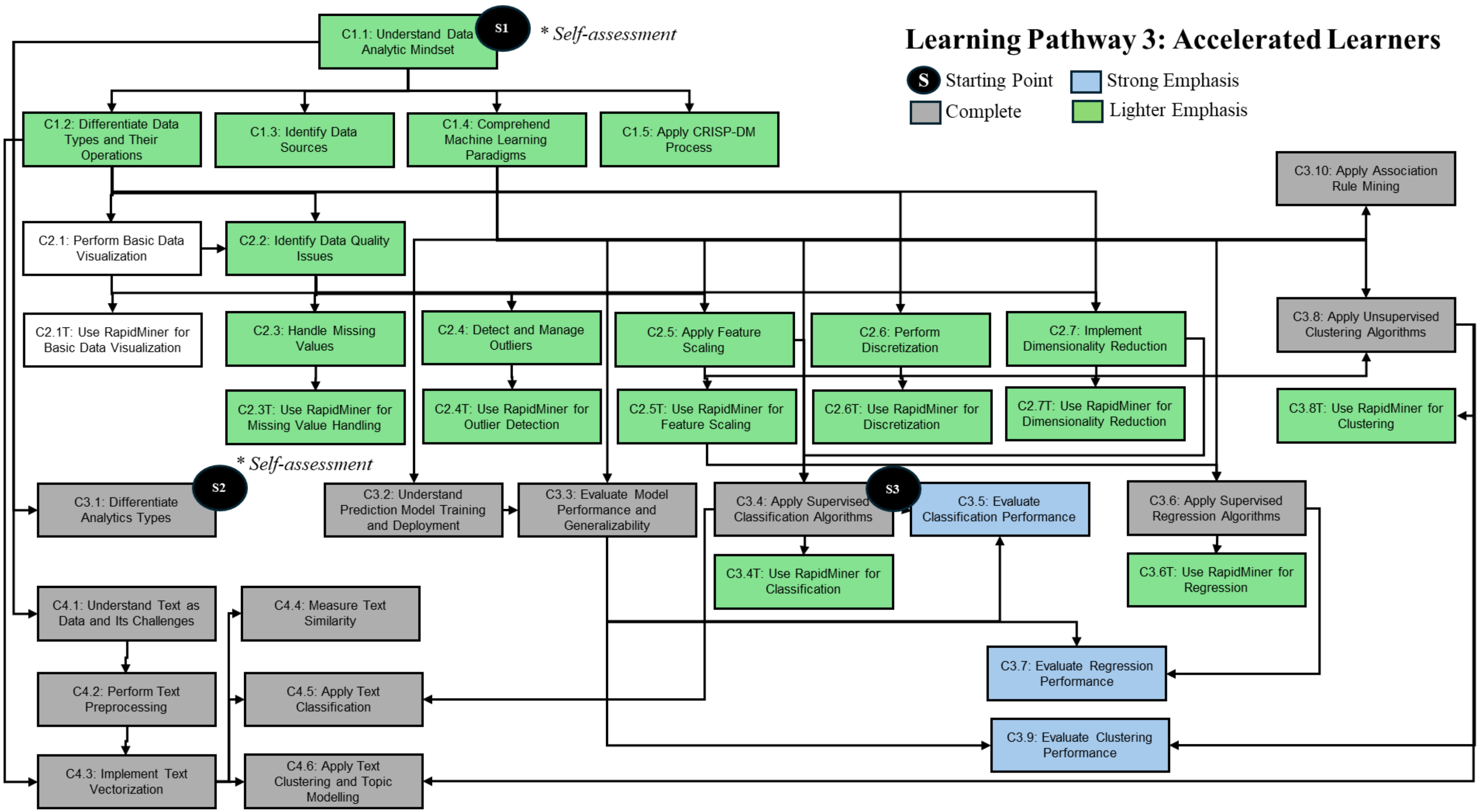

Figure 4: Visual representation of Learning Pathway 3 - Accelerated Learners