Deliverable 1.1 – List of Learning Needs

Introduction

The EduVerse intervention has been used in two iterations of IM0503 Data Analytics course, which is a part of the Business Process Management and Information Technologies (BPMIT) master's program. The number of students who took the course were N=66 in 2023/24 and N = 51 in 2024/25. Since BPMIT is an interdisciplinary program, it attracts students from various educational backgrounds. Based on the program enrolment records and the teacher's observations in the course, we deduced two major categories of educational background as **Category 1: Business Background** and **Category 2: Technical Background**. In addition to these broad educational background categories, we identified a category that covers a significantly different learning behavior independent from the education background, and we decided to include this as a third category; **Category 3: Accelerated Learners**.

Each category has a difference in their learning needs. This deliverable includes the learning needs of these three categories of learners, compiled based on the conversations with the Student User Committee and Teacher User Committee members.

The learning needs elaborated in this deliverable are used to develop the competency framework and adaptive learning pathways in the deliverable **D1.2 Alternative Learning Paths per Learning Unit**. These learning needs coupled with the competency framework and adaptive learning pathways can be used by other education professionals who aim at creating an adaptive learning experience for their interdisciplinary courses to facilitate learners from a diverse set of backgrounds.

Learning Needs:

Category 1: Business Background

Audience: This list of needs is tailored for individuals with a non-technical background, such as those in business administration, management, marketing, or strategic planning roles. They are typically comfortable with high-level concepts, strategic frameworks, and decision-making processes. Their prior exposure to advanced statistics, programming, or machine learning algorithms is likely limited or non-existent. Their primary objective is to use data analytics for strategic insights, effectively communicate with technical data teams, and make informed, data-driven business decisions.

Learning Needs & Justification:

1. **Conceptual Understanding of ML, AI, and Analytics Types**
    - **Need:** To gain a strong, intuitive grasp of *what* machine learning and artificial intelligence are, their core capabilities, and the distinctions between supervised, unsupervised, semi-supervised, and reinforcement learning paradigms. Additionally, they need to comprehend the progression from descriptive to predictive and prescriptive analytics and the "Data to Wisdom" hierarchy.
    - **Justification:** These learners are not expected to build detailed models themselves, but they must understand *what is possible* with data analytics to identify relevant business problems that can be solved using these methods. This foundational knowledge is crucial for articulating business needs to technical teams, evaluating proposed solutions, and interpreting the strategic implications of analytical results. Without it, they risk misidentifying opportunities, commissioning inappropriate projects, or misinterpreting outcomes.
    - **Pedagogical Needs:** Learning materials should heavily rely on **real-world business case studies, analogies, and high-level overviews**. Technical jargon should be minimized or clearly

explained. **Interactive discussions** focusing on strategic applications, ethical considerations, and business value are essential, rather than mathematical derivations or coding exercises.

2. **CRISP-DM as a Project Management Framework**
   - **Need:** A clear understanding of the Cross-Industry Standard Process for Data Mining (CRISP-DM) as a structured, iterative framework for managing data analytics projects, from initial business understanding to final deployment.
   - **Justification:** As managers or project sponsors, they will be responsible for overseeing or initiating data analytics initiatives. CRISP-DM provides a common language and systematic approach for planning, monitoring progress, and ensuring that technical efforts remain aligned with overarching business objectives. Understanding the *flow* and *interdependencies* between phases (e.g., how business understanding shapes data preparation) is vital for effective project governance.
   - **Pedagogical Needs:** Focus on the project lifecycle, typical roles involved in each stage, and critical decision points. **Group exercises simulating project planning or review meetings** around a data analytics project can be highly beneficial.

3. **Data Quality Awareness & Business Impact**
   - **Need:** To recognize common data quality issues (e.g., missing values, outliers, inconsistencies, noise) at a conceptual level and, more importantly, to understand the *severe consequences* of poor data quality on analytical reliability, model performance, and subsequently, on business decisions.
   - **Justification:** While they may not perform data cleaning, business-background learners must appreciate the effort, time, and resources required for robust data preparation. This awareness will inform data collection strategies, budget allocation for data engineering, and foster a healthy scepticism towards insights derived from potentially compromised data. It is about knowing *when to ask* about data quality and *what questions to ask*.
   - **Pedagogical Needs:** Visual examples of how data quality issues can lead to misleading or erroneous conclusions. **Discussions centred on the tangible business costs and risks** associated with inaccurate or incomplete data.

4. **Interpretation of Data Visualizations**
   - **Need:** To be proficient in effectively reading, interpreting, and drawing meaningful conclusions from various data visualizations, including line charts, scatter plots, histograms, pie charts, box plots, and bubble charts.
   - **Justification:** Data visualization is the primary language through which technical teams communicate their findings. If business strategists cannot accurately interpret these visuals, the insights generated by data analysts will not translate into actionable business intelligence. It's about discerning trends, identifying anomalies, and understanding distributions.
   - **Pedagogical Needs:** Extensive exposure to diverse visualizations, accompanied by clear explanations of what each chart conveys in a business context. **Interactive exercises focused on interpreting charts, identifying key takeaways, and formulating questions** based on the visual information.

5. **Evaluation Metrics for Business Objectives & Trade-offs**
   - **Need:** A profound understanding of what various model performance metrics (e.g., Precision, Recall, F1-Score, Expected Value for classification; MAE, MSE, RMSE for regression; Silhouette, Dunn for clustering) *mean in terms of business impact and risk*. Crucially, they need to grasp the *trade-offs* between these metrics and how to use them for making model selection decisions aligned with specific business costs and benefits.

- **Justification:** Model accuracy, while important, is often not the sole determinant for business decisions. A model with slightly lower overall accuracy but significantly fewer "false negatives" (e.g., in fraud detection, where each false negative is a direct loss) might be strategically superior. Business-background learners need to guide technical teams on which metrics are most critical given the specific business context and risk appetite.
- **Pedagogical Needs: Scenario-based learning where different models with varying performance profiles are presented**, and learners must select the "best" model by weighing business objectives, costs, and benefits. Emphasis on the *implications* of each metric rather than its mathematical calculation.

6. **Understanding Algorithm Applications & Limitations**
    - **Need:** To know *when and where to apply* various data analytics tasks (classification for predicting categories, regression for predicting continuous values, clustering for segmentation, association rule mining for patterns, text classification for sentiment, topic modeling for themes) to address specific business challenges.
    - **Justification:** This knowledge empowers them to identify suitable opportunities for implementing data analytics solutions within their organizations, scope projects realistically, and critically assess the feasibility of proposed solutions. They need to ask, "Which type of problem does this algorithm best solve?" and "What are its limitations for our business case?".
    - **Pedagogical Needs: Business problem-to-algorithm mapping exercises**, discussions on industry examples for each algorithm type, and a focus on the *inputs, outputs, and interpretability* of models rather than internal mechanics.

Category 2: Technical Background

Audience: This list of needs is for students with a robust foundation in technical disciplines such as computer science, information science, engineering, mathematics, or statistics. They are generally comfortable with programming, logical problem-solving, and abstract mathematical concepts. They may have prior experience with basic algorithms, data structures, and perhaps some introductory machine learning. Their goal is to acquire the practical skills to build, implement, optimize, and troubleshoot data analytics solutions.

Learning Needs & Justification:

1. **Deep Mathematical & Statistical Foundations**
    - **Need:** A thorough understanding of the underlying mathematical principles (e.g., linear algebra, basic calculus for optimization, probability, set theory) and statistical concepts (e.g., descriptive statistics, inferential statistics, hypothesis testing) that form the foundation of data analytics algorithms and evaluation metrics. This includes understanding the derivations and assumptions behind formulas like z-scores, IQR, various distance metrics (Euclidean, Manhattan), entropy, information gain, MAE, MSE, Precision, Recall, etc.
    - **Justification:** To effectively implement, debug, optimize, and potentially innovate in data analytics, a technical-background learner needs to understand *how* algorithms fundamentally work, not just *what* they do. This deep understanding enables intelligent parameter tuning, informed algorithm selection, effective troubleshooting of model issues, and the ability to adapt solutions to novel problems.
    - **Pedagogical Needs:** Detailed explanations of formulas, theoretical underpinnings, and derivations. **Problem sets requiring calculations, manual algorithm tracing, and critical analysis of assumptions.**

2. **Proficiency in Data Preparation Techniques**
    - **Need:** Hands-on mastery of various data cleaning, transformation, and reduction techniques, including handling missing values, detecting and managing outliers, applying feature scaling (normalization, standardization), performing discretization, and implementing dimensionality reduction (feature selection, PCA). This encompasses knowing *when* to apply each technique, *how* to implement it correctly, and understanding its *precise impact* on the data and subsequent modeling.
    - **Justification:** Real-world datasets are inherently messy and require significant preprocessing. Technical implementers spend a substantial portion of their time on data preparation. Proficiency here is critical for ensuring data quality, preventing model biases, and preparing data for optimal algorithm performance.
    - **Pedagogical Needs: Extensive hands-on labs and exercises** for each technique. Focus on practical implementation details, common pitfalls, selection of appropriate parameters, and best practices.
3. **Algorithm Mechanics and Implementation**
    - **Need:** An in-depth understanding of the internal mechanisms and specific steps of various machine learning algorithms (e.g., Decision Trees, K-Nearest Neighbors, Linear Regression, Apriori, FP-Growth, K-means, DBSCAN, and their application to text data). This includes comprehending their core logic, input requirements, output interpretation, parameter settings, computational complexity, and the process of model training and prediction. They must be able to translate these concepts into practical implementations.
    - **Justification:** This knowledge is paramount for building robust and effective models, selecting the most appropriate algorithm for a given problem, and proficiently tuning model hyperparameters for optimal performance and interpretability. They need to be able to *reason about algorithm behavior* and *optimize their application*.
    - **Pedagogical Needs: Step-by-step walkthroughs of algorithm logic, pseudocode, and detailed workflows with explanations of each operator's role and configuration.** Challenging implementation exercises where they troubleshoot and optimize models.
4. **Rigorous Model Evaluation & Validation**
    - **Need:** To master different validation strategies (e.g., train-test splitting, k-fold cross-validation) and a comprehensive suite of performance metrics (e.g., Confusion Matrix, Precision, Recall, F1-Score, AUC-ROC, Expected Value for classification; MAE, MSE, RMSE, MAD for regression; Silhouette Analysis, Dunn Index, Davies-Bouldin Index for clustering). This entails understanding their mathematical definitions, how to compute them (both manually and via tools), and how to interpret them *accurately* to assess model performance and generalizability.
    - **Justification:** Rigorous evaluation ensures that models are robust, reliable, and will perform effectively on unseen data. Technical-background learners must be adept at identifying and mitigating issues like overfitting and underfitting, and selecting models based on sound statistical and performance criteria.
    - **Pedagogical Needs: Exercises in calculating metrics manually** and interpreting complex performance reports. Discussions on biases in evaluation, statistical significance, and advanced model validation best practices.
5. **Text Processing and Vectorization Mechanics**
    - **Need:** Hands-on ability to execute all text preprocessing steps (tokenization, stemming, lemmatization, stopword removal, text normalization) and implement various text vectorization techniques (e.g., One-hot encoding, Bag-of-Words with binary, term count, and term frequency

features, TF-IDF, N-grams). They need a clear understanding of *how* these techniques transform text data and their subsequent impact on feature representation and model performance.

- **Justification:** Text data, being unstructured, requires specialized processing to be amenable to machine learning algorithms. Technical implementers need to master these techniques to effectively transform raw text into a numerical format suitable for analysis and modeling.
- **Pedagogical Needs: Labs focused on building text processing pipelines**, comparing the outcomes of different techniques, and understanding the parameters involved. Practical exercises on calculating TF-IDF or N-gram features manually and with the tool.

6. **Troubleshooting & Debugging Analytical Workflows**
   - **Need:** To develop strong problem-solving skills to identify, diagnose, and resolve issues that arise during any stage of a data analytics project; from data ingestion and preprocessing to model training, evaluation, and deployment.
   - **Justification:** Real-world data analytics projects are complex and frequently encounter unexpected challenges. The ability to effectively troubleshoot why a model is underperforming, why data transformations are yielding unexpected results, or why an algorithm isn't converging is a critical skill for any implementer.
   - **Pedagogical Needs: Problem-solving scenarios, guided debugging exercises, and opportunities for independent troubleshooting** or conceptual problems.

Category 3: The Accelerated Learner

Audience: This list of needs relates to students who already possess a foundational understanding of machine learning and basic data preparation concepts, perhaps from an undergraduate course, extensive online learning (MOOCs), or initial professional experience. They are often proficient in general data concepts and may understand the *why* but need to solidify the *how*, particularly with specific tools and in-depth application of techniques covered in this course. They are motivated to fill knowledge gaps, gain practical tool proficiency, and deepen their understanding of advanced or specialized topics.

Learning Needs & Justification:

1. **Efficient Gap Identification & Targeted Learning**
   - **Need:** The ability to accurately self-assess their existing knowledge against the specific competencies of this course and efficiently identify precise areas where they need to deepen understanding or gain practical skills.
   - **Justification:** This allows them to avoid redundant learning, maximize their learning efficiency, and focus their efforts on new, more complex, or tool-specific topics. They need to validate their existing knowledge quickly and move on.
   - **Pedagogical Needs:** A clear and comprehensive **competency map with integrated self-assessment tools** (e.g., clear learning objectives, pre-quizzes, challenge problems, checklists) for each module. The course structure should be modular, allowing learners to easily navigate to specific units or topics without being forced through already mastered content.
2. **RapidMiner Tool Proficiency for Specific Implementations**
   - **Need:** To proficiently translate their existing theoretical understanding of data preparation and machine learning algorithms into practical, hands-on implementation, workflows, and parameter configurations.

- **Justification:** While they may understand the theory, using a new, powerful tool effectively requires dedicated practice and understanding its unique functionalities, conventions, and nuances. Bridging the gap between theory and specific tool application is key.
- **Pedagogical Needs: Concise, hands-on tutorials and labs focused on implementation** for each data analytics task, emphasizing the "how-to" and specific operator usage. "Challenge" exercises that require applying a learned algorithm in RapidMiner to new, unseen datasets.

3. **Nuances of Algorithm Selection, Tuning, and Performance**
    - **Need:** To move beyond basic algorithm understanding to a more refined appreciation of the subtle differences between similar algorithms, their specific assumptions, practical strengths and weaknesses, and how to effectively tune their hyperparameters for optimal performance in diverse scenarios. This includes understanding edge cases and when one algorithm is preferable over another.
    - **Justification:** Accelerated learners often have a surface-level understanding. This course should push them to a deeper, more practical understanding of advanced considerations in model building, including balancing performance, interpretability, and computational cost.
    - **Pedagogical Needs: Comparative analyses of similar algorithms, discussions on advanced hyperparameter tuning strategies, and case studies where algorithm selection and fine-tuning have critical implications** for model success.
4. **Advanced Evaluation Strategies & Business Alignment**
    - **Need:** A deeper dive into the sophisticated interpretation of performance metrics, understanding the inherent trade-offs (e.g., between Precision and Recall), and how to apply more complex evaluation schemes (like Expected Value) to align model performance directly with specific business objectives and risk assessments.
    - **Justification:** They likely understand basic metrics like accuracy but need to develop a more sophisticated and business-centric view of model evaluation that considers real-world impact, financial costs of errors, and strategic value.
    - **Pedagogical Needs: Advanced case studies with complex evaluation scenarios**, opportunities to critically analyze performance reports, and discussions on the ethical implications of different types of model errors in various business contexts.
5. **Integration of Specialized Topics**
    - **Need:** To integrate their existing general machine learning knowledge with specialized domains, particularly text analytics. This includes grasping the unique preprocessing and vectorization challenges and specific solutions for unstructured text data.
    - **Justification:** Text analytics represents a distinct application area within data science. Accelerated learners need to understand how general ML principles adapt to this specific data type, including the specialized techniques required for text (e.g., TF-IDF, N-grams, stemming vs. lemmatization) and how these choices impact downstream models.
    - **Pedagogical Needs:** Comprehensive modules on text analytics, emphasizing how the techniques differ from numerical data processing while highlighting where core ML principles still apply. Hands-on exercises for building full text analysis pipelines.
6. **Learning Efficient Workflows and Best Practices**
    - **Need:** To acquire insights into efficient data analytics workflows, common industry best practices, and practical strategies for avoiding common pitfalls throughout a project lifecycle.

- **Justification:** Even experienced learners can benefit significantly from structured guidance on efficient methodologies, project organization, and lessons learned from real-world projects, optimizing their future work.
- **Pedagogical Needs:** Dedicated sections on "tips and tricks," discussions on common mistakes made by practitioners, and practical examples of applying iterative, agile approaches (like CRISP-DM) to real projects for maximum efficiency.