LONDON METROPOLITAN UNIVERSITY

informatics college • pokhara

**Module Code & Module Title**

**CU6051NP Artificial Intelligence**

**25% Individual Coursework**

**Submission: Proposal**

**Academic Semester: Autumn Semester 2025**

**Credit: 15 credit semester long module**


**Student Name:** Sharon Gurung

**London Met ID:** 23048933

**College ID:** NP04CP4A230114


**Assignment Due Date:** 17/12/2025.

**Assignment Submission Date: 17/12/2025**

**Submitted To:** Jeevan Prakash Pant

| | |
|---|---|
| **GitHub Link** | *https://github.com/Norahs-00/Learning_recommendation_system.git* |

*I confirm that I understand my coursework needs to be submitted via MST under the relevant module page before the deadline in order for my coursework's milestone to be accepted and marked. I am fully aware that late submissions will be treated as non-submission and a mark of zero will be awarded.*

## Table of Contents

## Table of Figures

## Table of Tables

# 1. Introduction

## 1.1 Explanation of the Topic and AI Concepts Used

AI has become increasingly used in the field of education to design intelligent systems that improve learning experiences and support informed decision-making. A major use of AI in the education sector is recommendation systems, which process large volumes of data and filter information to reduce information overload. While recommendation systems are commonly used in areas like e-commerce and entertainment to suggest products, movies, or music, their application in education has allowed for more personalized learning experiences by guiding learners to relevant courses and learning content according to their individual characteristics, interests, and academic needs (Tilahun & Sekeroglu, 2020).

Modern recommendation systems are based on Machine Learning (ML), a subset of AI that enables a system to learn patterns from historical data and continually refine recommendations without explicit programming. Supervised learning is used where historical data contains known outcomes, such as course ratings, enrolments or learner preferences. In this project, these methods such as Logistic Regression, Decision Trees, and Random Forests are explored to predict whether to recommend a course based on known outcomes such as learner preferences or engagement, logistic regression as a baseline model for binary classification, decision trees for rule-based recommendations, and random forests to increase accuracy, reduce overfitting. In addition, unsupervised learning methods like clustering are used to group learners with similar interests or behaviours so the system can recommend courses to entire learner groups. Hybrid recommendation approaches are considered by combining supervised and unsupervised methods to enhance the performance and robustness (Ren, et al., 2022).

The application of these AI and machine learning techniques brings several key benefits to educational recommendation systems:

1. Personalization: Learners receive course suggestions based on their interests, prior knowledge, learning goals and career goals.
2. Efficiency: Learners spend less time searching through thousands of available courses according to their needs and preferences.
3. Engagement: This increases learner engagement with individual goals and their interests. It recommends courses that are better fit with learner preferences which increases motivation and completion rates.
4. Scalability: Systems can manage large datasets and provide recommendations to millions of learners at the same time.

In this project, AI techniques will be used to develop a Learning Recommendation System based on the Coursera-course-dataset, which contains course title, enrolment numbers, ratings, reviews, instructor names, and organisations. The goal is to develop a system that uses supervised, unsupervised, and hybrid machine learning techniques to generate recommendations for specific learners.

Therefore, supervised, unsupervised and hybrid machine learning methods allow the proposed learning recommendation system to improve accuracy, personalisation and scalability. Each method addresses different problems within online learning platforms and their use enhances the ability of the system to deliver relevant and meaningful course recommendations to various learners.

## 1.2 Explanation of the Chosen Problem Domain

Online learning platforms such as Coursera, Udemy, and EdX have transformed access to education by offering thousands of courses across a wide range of subjects, including computer science, engineering, business, medical sciences and the arts. However, learners may experience difficulty in identifying courses that are most relevant to their personal goals, prior knowledge, and career aspirations. This challenge is commonly called information overload which can lead to confusion, poor decision-making, and reduced learner engagement (Ren, et al., 2022).

Despite the rapid growth of online learning platforms, quantitative data shows that learner engagement and the completion rate of the course remain low. Studies of Massive Open Online Courses (MOOCs) report that the completion rate between 5 % and 15% with median of approximately 12.5% which means nearly 85-90%of learners do not complete their courses (Peterson, 2013). Over 50% of learners who register for online courses never start the course which reflects significant learner loss at early stage. The scale of this problem is concerning for online learning systems. The scale of online education by 2021 had reached approximately 220 million learners worldwide with around 40 million new learners enrolled in that year (Shah, 2021). The majority of recommendation systems on online learning platforms primarily focus on popularity-based criteria like number of enrolments or average course ratings. While these metrics offer a sense of course quality, they fail to reflect the unique preferences of individual learners or acknowledge that learners from diverse cultural or professional backgrounds might prioritize different course features, such as hands-on application or theoretical complexity. Consequently, this can lead to misaligned course enrolments, decreased completion rates, wasted effort and diminished learner satisfaction (Tilahun & Sekeroglu, 2020).

Therefore, this project focuses on the development of a personalized course recommendation system. Personalization in education involves adapting learning experiences to individual learner characteristics such as prior knowledge, interests, learning styles, and career paths. The

Coursera-course 2024 dataset that consists of 6,645 records with attributes such as course title, enrolment number, rating, review, instructor name, and organisation is utilized. The goal of this project is to develop an AI system that can recommend courses more closely aligned with learner profiles.

This system directly addresses major issues in online learning. It reduces information overload by filtering and ranking relevant courses, supports learner diversity through customized recommendations and enhances retention and achievement by matching course complexity and material with learner capabilities. By promoting sustained engagement, the system can enhance educational results and overall satisfaction for learners. In summary, the chosen problem domain highlights the critical need for AI-driven personalization in online education. By moving beyond traditional popularity-based recommendation methods and incorporating machine learning techniques, the proposed system aims to enhance course selection, promote learner success, and support equitable access to education.

## 1.3 Aims and Objectives

The main aim of this project is to develop and assess a learning course recommendation system that helps learners choose courses from extensive online platforms, like Coursera. This system will utilize machine learning techniques. Course metadata provide suggestions that match learner preferences, existing knowledge and professional objectives. It aims to minimize information overload, enhance learner engagement, and increase course completion rates.

For accomplishing this goal, the resulting objectives have been established:

- To analyse an online course dataset using exploratory data analysis (EDA) and understand key attributes related to course recommendation.
- To apply supervised machine learning methods, including Logistic Regression, Decision Trees, and Random Forests, for predicting suitability of the course.
- To utilize unsupervised learning methods such as clustering to group courses or learners with similar interests.
- To develop a hybrid recommendation method that combines multiple methods to improve accuracy and scalability.
- To evaluate the effectiveness of the system using appropriate performance metrics of the proposed system in generating relevant and meaningful course recommendations.

## 2. Background

### 2.1 Research Work Done

Research on AI-driven personalized course recommendation systems has evolved progressively, with each phase refining earlier limitations and expanding system intelligence. Early foundational work was focused on structured decision-making for academic planning, while later studies introduced contextual Modeling, deep learning, and real-time personalization.

#### 2.1.1 Research 1: Artificial Intelligence in Adaptive Education

The groundwork for personalized course recommendation was established by (Xu, et al., 2016), who addressed the challenge of recommending not just individual courses but entire course sequences. Their work framed course recommendation as a sequential decision-making problem constrained by prerequisites, course availability, and learner background. By integrating dynamic programming with contextual multi-armed bandits, the authors demonstrated that adaptive algorithms could learn from historical student performance to optimize outcomes such as graduation time and academic success. This study marked a critical shift from static advising toward data-driven, adaptive recommendation policies, laying the algorithmic foundation for later personalization research.

#### 2.1.2 Research 2: AI for Personalized Learning in Higher Education

Building on this algorithmic perspective, (Tilahun & Sekeroglu, 2020) advanced course recommendation systems by introducing an intelligent course advising model grounded in expert systems and rule-based reasoning. Unlike earlier purely algorithmic approaches, their model incorporated institutional regulations, curriculum structures, and student academic history into a unified advisory framework. This refinement addressed practical deployment challenges in higher education by ensuring flexibility across departments while maintaining persistent academic data. The study extended earlier work by embedding AI recommendations within real institutional constraints, thereby improving usability and advisor support.

#### 2.1.3 Research 3: Multi-Model Course Recommendation Framework

As personalization demands increased, attention shifted toward understanding contextual dependencies among courses. (Islam & Hosen , 2022) refined prior sequence-based approaches by explicitly modeling prerequisite relationships as contextual signals in recommendation systems. Their work demonstrated that ignoring prerequisite structures could lead to suboptimal or infeasible recommendations. By incorporating prerequisite context into the recommendation process, the study improved both accuracy and pedagogical validity. This contribution strengthened earlier sequence-based models by aligning algorithmic recommendations with learning progression logic, ensuring that suggested courses supported coherent skill development.

The evolution toward richer learner modeling continued with (Ren, et al., 2022), who introduced a deep learning-based multimodal course recommendation framework. Moving beyond structured academic records alone, their approach integrated diverse data modalities such as course videos, textual descriptions, and user interaction behaviour using LSTM networks with attention mechanisms. This represented a significant refinement over earlier models by capturing both explicit and implicit learner preferences. The multimodal design improved recommendation accuracy and addressed information overload in large-scale online learning environments, signalling a shift toward behaviour-aware and content-rich personalization.

### 2.1.4 Research 4: AI for Lifelong Learning

Most recently, (Khan & Polyzou, 2024) extended this path by focusing on session-based course recommendation, emphasizing short-term learner intent rather than long-term historical profiles. Their work acknowledged that learner interests evolve dynamically within browsing sessions and proposed models capable of adapting recommendations in real time. This refinement addressed cold-start and sparsity issues inherent in traditional collaborative filtering approaches. By prioritizing temporal interaction patterns, their study represents the current state of the art, where course recommendation systems are not only personalized but also responsive, adaptive, and context-aware. Together, these studies demonstrate a clear horizontal progression: from structured sequence optimization (2016), to institutional intelligence (2020), contextual prerequisite modeling (2022), multimodal deep learning (2022), and finally real-time session-aware personalization (2024).

## 2.2 Review and Analysis of Existing Work

### 2.2.1 Review and Analysis of Research 1

The work by Xu et al. (2016) introduced a personalized course sequence recommendation framework that models learner progression over time. A major strength of this research is its ability to incorporate prerequisite constraints and long-term academic objectives, enabling recommendations that support structured learning pathways. This makes it particularly effective in traditional university environments where course dependencies are well defined. However, a key limitation is its reliance on rigid academic structures and sequential course data, which reduces its suitability for open online platforms such as Coursera. In such platforms, learners often follow non-linear learning paths and pursue diverse goals, making strict sequence-based recommendations less practical.

### 2.2.2 Review and Analysis of Research 2

Tilahun and Sekeroglu (2020) proposed an intelligent academic advising system using expert systems and rule-based techniques. One of the strengths of this approach lies in its transparency and interpretability, as recommendations are generated through clearly defined rules that can be easily understood by both learners and educators. Additionally, the system aligns well with institutional academic policies. However, the rule-based nature of the system limits its scalability and adaptability. It struggles to handle large and diverse datasets and cannot easily adjust to evolving learner preferences, which is a significant drawback for massive online learning environments.

### 2.2.3 Review and Analysis of Research 3

The study by Islam and Hosen (2022) presented a multi-model machine learning framework for course recommendation that integrates multiple predictive models to improve accuracy. A key strength of this research is its ability to handle complex academic constraints while achieving higher predictive performance compared to single-model approaches. This demonstrates the potential of hybrid systems in improving recommendation quality. Nevertheless, the framework relies heavily on structured academic datasets and lacks extensive validation using real-world learner interaction data. This limits its effectiveness in dynamic platforms where learner behavior is inconsistent and data sparsity is common. Ren et al. (2022) advanced the field by introducing a deep learning-based multimodal recommendation system that incorporates textual, behavioral, and contextual data. The primary strength of this approach is its ability to capture rich learner preferences and significantly reduce information overload by producing highly personalized recommendations. However, the model's complexity presents several limitations. Deep learning systems require substantial computational resources and large volumes of data, making them challenging to deploy in resource-constrained educational settings. Additionally, the lack of explainability in such models can reduce user trust and limit their adoption in educational contexts.

### 2.2.4 Review and Analysis of Research 4

More recently, Khan and Polyzou (2024) explored session-based recommendation systems that focus on learners' short-term intentions. The strength of this approach lies in its ability to address cold-start problems and rapidly adapt to learners with limited historical data, which is particularly relevant for online platforms. However, this approach prioritizes immediate learner behavior and may overlook long-term learning goals and knowledge progression. Furthermore, continuous session tracking requires advanced infrastructure, which may pose scalability and privacy challenges.

In summary, existing research demonstrates significant progress in improving personalization and recommendation accuracy in educational systems. However, limitations related to scalability, interpretability, data dependency, and adaptability remain unresolved. These gaps justify the need for a balanced learning recommendation system that integrates supervised, unsupervised, and hybrid machine learning techniques using readily available course metadata. The proposed system aims to leverage these strengths while addressing the identified limitations to provide scalable, transparent, and effective course recommendations.

## 2.3 Analytical Review of Existing Systems on the Problem Domain

In recent years, large-scale online learning platforms have developed a number of recommendation mechanisms to help learners choose courses, ranging from collaborative filtering and popularity-based ranking, in which recommendations were based on enrolment counts, average ratings, and simple learner–course interaction matrices, to more complex approaches that take into consideration both learner and course features (Ziegler, et al., 2017).

As online learning platforms expanded, hybrid recommendation systems emerged that combined content-based filtering with collaborative signals to consider both course information and user interaction patterns, these systems improved the relevance of recommendations (Buitrago & Chiappe, 2019). With the rise of MOOC(Massive Open Online Courses), researchers have also investigated knowledge-aware and constraint-based systems to enhance academic consistency, but systems that incorporate curriculum constraints and learner background information have shown to improve the validity of recommendations requiring structured institutional data that is typically not available on open platforms like Coursera (Tilahun & Sekeroglu, 2020). More recent systems introduce fairness-aware and explainable recommendation models to address ethical and transparency concerns, but these systems tend to sacrifice predictive power for transparency (Liu, et al., 2020).

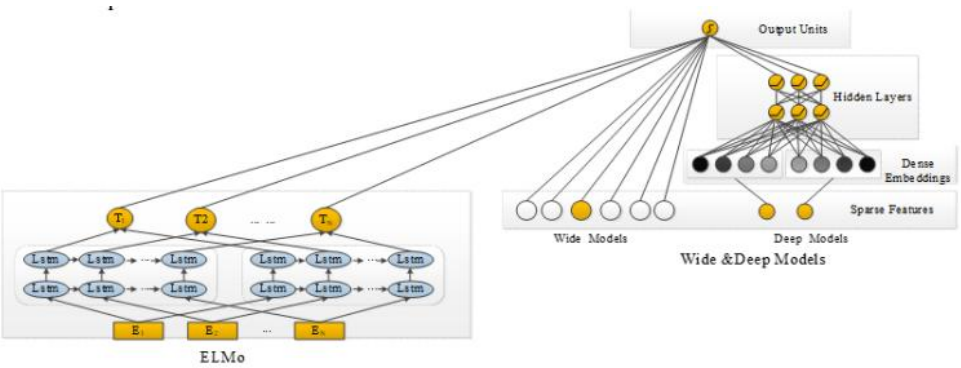


*Figure 1: Online Learning Resource Recommendation Method Based on Wide & Deep and Elmo Model (Liu, et al., 2020)*

State of the learning platforms are evolving to context-aware and session-based systems that can dynamically adapt recommendations based on short-term learner behaviour. However, these systems face challenges in scalability, data privacy, and real-time infrastructure needs (George & La, 2024). In conclusion, existing learning recommendation systems have evolved from static popularity-based models to adaptive and context-aware frameworks, but they are still limited by personalization depth, fairness, explainability, and scalability, justifying the need for a balanced system that can leverage structured course metadata while remaining scalable and interpretable.

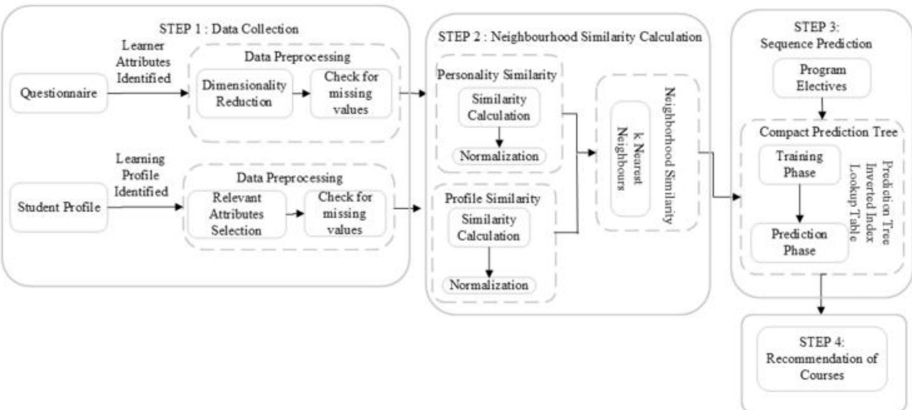


*Figure 2: Course recommendation framework (George & La, 2024)*

## 2.4 Dataset Description

The dataset used for this project is the Coursera-course 2024 dataset that consists of 6,645 records in CSV format. It includes attributes such as course title, enrolment numbers, average ratings, number of reviews, instructor names, and organizations. This dataset provides rich metadata for building learner profiles and generating recommendations. Enrolment and rating features help identify high-quality and widely accepted courses, while instructor and organization information enables similarity-based filtering. However, the absence of explicit learner interaction fuels limit of the application of pure collaborative filtering methods, supporting the need for hybrid recommendation approaches. The dataset includes the following features:

| **Feature Name** | **Description** | **Data Type** |
| --- | --- | --- |
| course_id | Unique identifier for each course | Integer |
| title | Course title | String |
| organization | Organization or university offering the course | String |
| instructor | Instructor name(s) | String |
| level | Course difficulty level | Categorical (String) |
| certificate_type | Type of certificate offered | Categorical (String) |
| enrolled | Number of learners enrolled | Integer |
| rating | Average course rating (out of 5) | Float |
| num_reviews | Number of learner reviews | Integer |
| duration_weeks | Estimated duration of the course in weeks | Integer |
| skills | Skills covered in the course | String |
| language | Language of instruction | String |
| url | Course webpage link | String |

*Table 1: Data Dictionary of the system*

# 3. Solution

## 3.1 Explanation of the Solution

This project proposes an AI-driven personalized learning course recommendation system that assist learners in selecting suitable courses from large-scale online learning platforms. The proposed solution addresses the limitations of traditional popularity-based recommendation methods by incorporating machine learning techniques that analyze course metadata and learner-related features to generate personalized recommendations. The system aims to reduce information overload, improve learner engagement, and support informed decision-making by tailoring course suggestions to learner preferences, prior knowledge indicators, and course characteristics.

The system is developed using the Coursera-course 2024 dataset, which contains structured metadata such as course titles, enrolment numbers, ratings, reviews, instructors, difficulty level and organizations. Since individual learner interaction data is limited, the proposed solution focuses on content-based and hybrid recommendation strategies, ensuring scalability and applicability in real-world online education platforms. Data preprocessing is performed to ensure the quality and consistency of the data. This includes handling missing values, removing duplicate records, normalizing numerical features and encoding categorical variables. These steps are important to improve reliability of model and reduce noise during training the data.

## 3.2 System Architecture and Workflow

The proposed learning recommendation system follows a modular architecture consisting of four main components: data preprocessing, feature engineering, recommendation engine, and output generation. Initially, the raw dataset undergoes preprocessing, which includes handling missing values, normalising numerical features such as number of enrolment and ratings, and encoding categorical variables such as instructor names and organizations. This step ensures that the dataset is suitable for machine learning models.

Next, feature engineering is performed to extract meaningful representations from the dataset. Course popularity, quality indicators, and organizational attributes are transformed into numerical vectors that capture course relevance. These features form the basis for training machine learning models. The core of the system is the recommendation engine, which applies multiple AI algorithms to generate ranked course recommendations. Supervised learning models predict the likelihood that a course should be recommended, unsupervised learning groups similar courses or learners, and hybrid methods combine the strengths of both methods. Finally, the system outputs a ranked list of recommended courses for learners, prioritising relevance, quality, and diversity. This workflow ensures that recommendations are both personalised and academically meaningful.

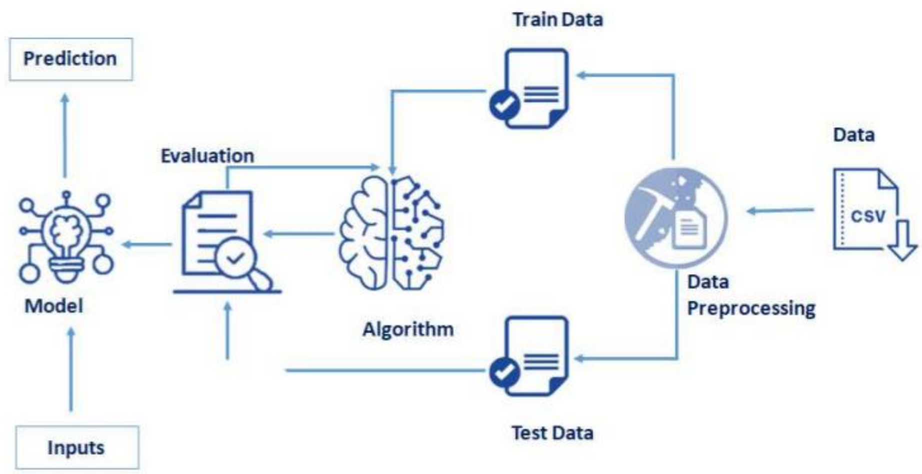


*Figure 3: System Architecture*

## 3.3 AI Algorithms Used

### 3.3.1 Supervised Learning Algorithms

Supervised learning techniques are used to predict whether a course should be recommended based on known outcomes such as ratings, enrolment levels, and review counts.

Logistic Regression is used as a baseline classifier due to its simplicity and interpretability. It estimates the probability that a course is relevant by modelling the relationship between course features and recommendation outcomes(0 or 1).

$$P(y = 1 \mid x) = 1 / (1 + e^{\wedge} - (\beta 0 + \beta 1x1 + \beta 2x2 + \ldots + \beta nxn))$$

Here, $x_1, x_2, \ldots, x_n$ represent course features such as rating, number of enrolments, number of reviews, course level and duration of the course. The parameters β0, β1,…βn are learned weights. The sigmoid function converts the linear combination into a probability between 0 and 1. This equation gives the probability that a course should be recommended. If the probability equals to 1 then the course is highly suitable. If it is close to 0 then the course is not suitable. A threshold (0.5) is used to convert this probability into a recommendation decision.

Decision Trees are used to capture non-linear relationships between course attributes and recommendation decisions.

$$Entropy(S) = -\sum_{\{i=1\}_{i}^{\{c\}p}\backslash log_2} p_i$$

$$IG(S,A) = Entropy(S) - \sum_{\{v\}\backslash frac\{|S_v|\}\{|S|\}Entropy(S_v)}$$

Here, S represents the complete course dataset, A denotes a course feature such as rating or number of enrolment and $Sv$ represents the subsets formed after splitting on feature A. Entropy measures the impurity of the dataset. Information Gain shows how much uncertainty is reduced when splitting data using a course feature such as rating or enrolment. Decision Trees helps generate rules for course recommendation such as recommending courses with higher ratings and enrolment levels.

Random Forests, an ensemble learning method, combine multiple decision trees to improve prediction accuracy and reduce overfitting. By combining the results of several trees trained on random subsets of the data, Random Forests provides more robust and accurate recommendations even when the dataset contains noise or imbalance.

$$\widehat{}\{y\} = majority\ vote\left(T_{1(x)}, T_{2(x)}, \dots, T_{n(x)}\right)$$

Here, represents the prediction made by the $t^{th}$decision tree using course features such as rating, enrolment count, and reviews. The final recommendation $\hat{y}$ is determined using majority voting across all the trees.

3.3.2 Unsupervised Learning Algorithms

Unsupervised learning techniques are applied to identify patterns in course data without relying on labelled outcomes. K-Means clustering group courses with similar characteristics based on ratings, enrolment patterns, and content-related features. It aims to minimize the within cluster sum of squares (WCSS) which represents total squared distance between each course and its assigned cluster centroid. This allows the system to recommend clusters of related courses to learners with similar interests. Unsupervised learning is particularly useful for handling cold-start scenarios, where limited information about learners is available.

$$J = \sum_{\{j=1\}}^{\{k\}\sum_{\{x_i \in C_j\}}^{2}\left\|x_i - c_j\right\|}$$

Where:

J = Total clustering error

k = Number of clusters

$x_i$ = A course(feature)

$C_j$ = Cluster j

$c_j$ = centroid (average course)

$||x_i - c_j||^2$ = squared distance

Here, $x_i$represents a course feature, $c_j$represents the centroid of cluster $j$, and $k$ is the number of clusters. The algorithm groups courses with similar characteristics such as ratings, enrolment levels, and duration by minimizing the objective function. These clusters are then used to recommend groups of similar courses to learners with comparable interests.

### 3.3.3 Evaluation Metrics

Precision measures the accuracy of the recommendation system by showing the recommended courses according to the preference of the learner. A high precision value indicates that the system rarely recommends unsuitable courses.

$$Precision = \frac{TP}{TP + FP}$$

Recall measures the ability of the system to identify if all relevant courses by showing relevant courses that are successfully recommended to the learner. A high recall value represents that the system does not miss suitable courses.

$$Recall = \frac{TP}{TP + FN}$$

F1 score is the combination of precision and recall that provides balance evaluation of the system's performance. It is useful in recommendation of highly relevant courses for learner.

$$F1 = 2 \times \frac{(Precision * Recall)}{(Precision + Recall)}$$

### 3.3.4 Hybrid Recommendation Approach

To overcome the limitations of individual methods, the proposed system adopts a hybrid recommendation approach that integrates supervised predictions with unsupervised clustering results. This approach enhances recommendation accuracy while maintaining flexibility. For example, supervised models rank courses based on predicted relevance, while clustering ensures diversity by selecting recommendations from multiple course groups. Hybridization also reduces bias toward highly popular courses and improves personalisation by balancing popularity, quality, and similarity. The final recommendation score is weighted combination of supervised prediction and clustering which ensures relevance and diversity.

$$Score_{\{final\}} = \alpha \cdot Score_{\{supervised\}} + (1 - \alpha) \cdot Score_{\{cluster\}}$$

## 3.4 Pseudocode of the solution

START

IMPORT required libraries

  pandas for dataset handling

  numpy for numerical operations

  sklearn for preprocessing and machine learning

    StandardScaler for normalization

    train_test_split for data splitting

    LogisticRegression, DecisionTreeClassifier, RandomForestClassifier

    KMeans for clustering

    accuracy_score, precision_score, recall_score

LOAD Coursera-course-dataset (CSV file) into DataFrame

DATA PREPROCESSING

  REMOVE duplicate records

  HANDLE missing values

  SELECT relevant features (ratings, enrollments, reviews, course category)

  ENCODE categorical features (instructor, organization)

  NORMALIZE numerical features using StandardScaler

SUPERVISED LEARNING

  DEFINE target variable (Recommend =1, Not Recommend=0)

  SPLIT dataset into training and testing sets (80% training, 20% testing)

  TRAIN Logistic Regression model

  TRAIN Decision Tree model

  TRAIN Random Forest model

  EVALUATE supervised models

```
      CALCULATE accuracy and precision scores

UNSUPERVISED LEARNING
     APPLY K-Means clustering on course features
     GROUP similar courses into clusters
      IDENTIFY course similarity patterns
HYBRID RECOMMENDATION
  FOR each course:
    CALCULATE supervised relevance score

    IF supervised relevance score < predefined threshold THEN
      USE cluster-based recommendation
    ELSE
      USE supervised ranking
    END IF

  COMBINE supervised rankings with clustering results
  RANK courses based on relevance, similarity, and diversity

FINAL RECOMMENDATION
  INPUT learner preferences
  MATCH learner preferences with ranked courses
  OUTPUT Top-N recommended courses

END
```

## 3.5 Diagrammatical representations of the solution

### 3.5.1 Flowchart

Flowchart is the graphical representation of step-by-step operational workflow events or actions to make better decision. It is the best way for the beginners to create a program for general purpose. Oval shape represents start and stops of the program, parallelogram represent the input and output of the data, arrows show the direction, rectangle represent the tasks or process and diamond for decisions.

It starts with loading Coursera course 2024 dataset then followed by data preprocessing and normalizing numerical attributes. It illustrates the use of supervised learning methods to predict known outcomes (course relevance) and unsupervised method including clustering to group similar courses. A decision shape determines whether supervised relevant scores that meet a defined threshold. If not then cluster based recommendation are applied. Finally, the system generates a list of recommended courses based on learner preferences, interest and learning goals. Flowcharts are commonly used in system design to clearly visualize algorithm, make complex workflow easier to understand and improve decision making (Charntaweekhun & Wangsiripitak, 2006).

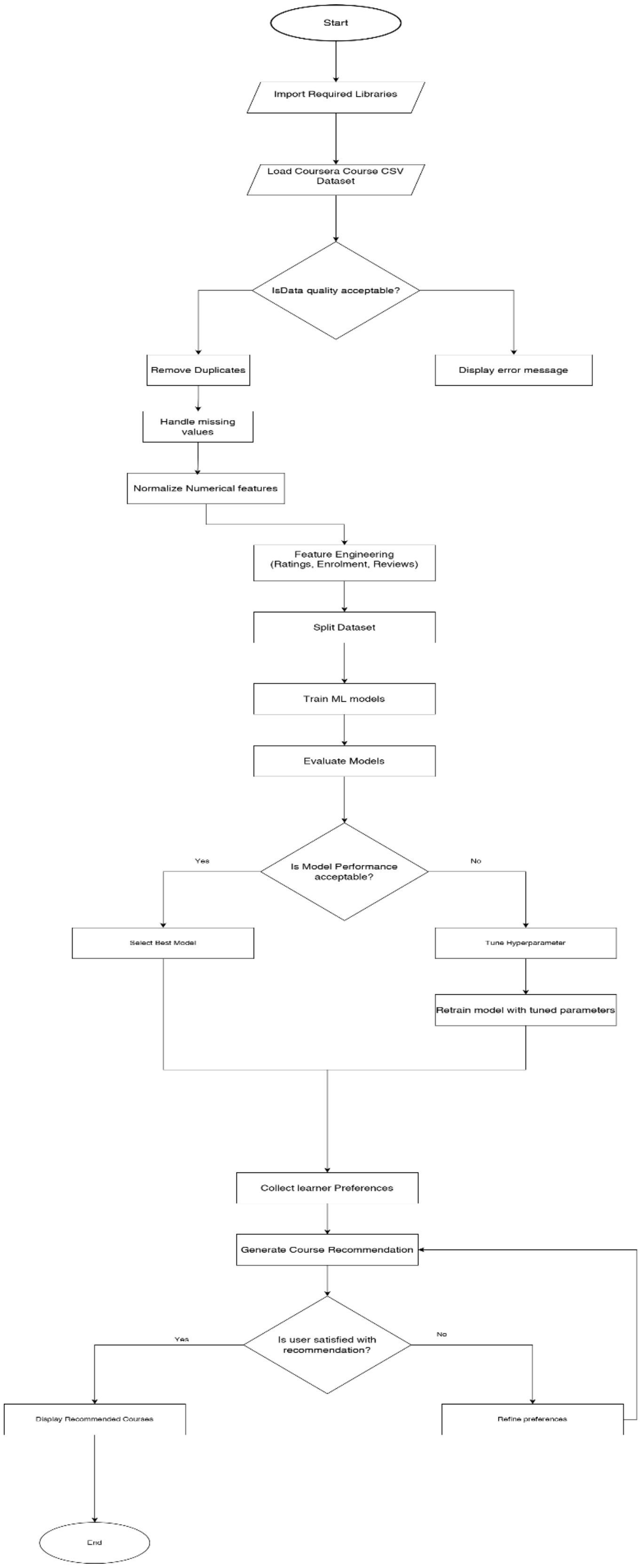


*Figure 4: Flowchart of the system*

### 3.5.2 State transition Diagram

A State Transition Diagram visually represents a finite state machine, utilized to model objects that have a limited number of states and their interactions with the external environment through state changes driven by events (ScienceDirect, 2012). It is made up of nodes that symbolize states and directed edges that show transitions marked with event names. The system transitions from the initial state to data loading, data preprocessing, model training, and recommendation generated states, before reaching the final output state where personalized course recommendations are delivered to learner.

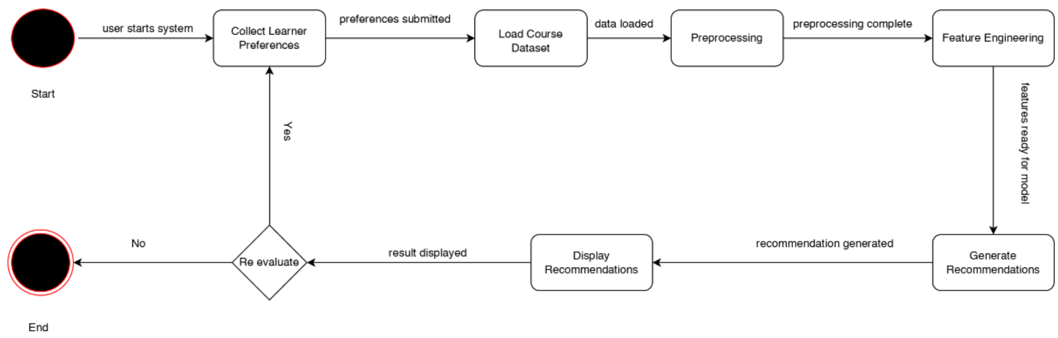


*Figure 5: State transition Diagram*

# 4. Conclusion

## 4.1 Analysis of the Work Done

This coursework has presented the design and conceptual solution for a learning course recommendation system using Artificial Intelligence (AI) to improve personalization in online education platforms such as Coursera, Udemy, and EdX. Through a study, how recommendation systems have evolved from simple rule-based models to advanced hybrid approaches that adapt to user sessions.

While these systems are highly accurate, they still struggle with scalability, interpretability, and cold-start problems. Course enrolment and ratings help identify popular, high-quality options, while instructor and organization details enable similarity-based recommendations. However, without detailed learner interaction data, pure collaborative filtering cannot be effectively applied, making hybrid approaches essential. The project analysed the Coursera-course-dataset (6,645 rows) and applied preprocessing techniques to ensure data quality. Exploratory Data Analysis (EDA) discovered important trends such as the correlation between ratings, reviews, and enrolments, which informed the choice of algorithms. Multiple ML models: Logistic Regression, Decision Tree, Random Forest, and Hybrid Filtering were proposed to generate recommendations. Evaluation metrics such as Precision, Recall, F1 Score were identified to measure system performance. By integrating these components, the system demonstrates how AI can transform traditional recommendation methods into personalized, scalable, and efficient solutions for online learning.

In context to real world, this system can support learners in making informed decisions, reduce course dropout rates, and support platforms in delivering personalized learning experiences. Future improvements may include incorporating real user interaction data, applying deep learning models for preference modeling. Overall, this project demonstrates how artificial intelligence can effectively address real educational challenges through well-designed recommendation systems.

## 4.2 Further Work

While the proposed system provides a strong foundation, several areas for improvement remain:

- Integration of Natural Language Processing (NLP): Incorporating learner feedback and reviews to refine recommendations.
- Multimodal Data: Using behavioural, emotional, and biometric data to enhance personalization.
- Explainable AI (XAI): Ensuring transparency in recommendations to build trust among learners and educators.
- Deployment: Developing a prototype application with a user-friendly interface and real-time recommendation capabilities.

Future iterations of the system can expand beyond Coursera to include multiple platforms, creating a unified recommendation engine for lifelong learning across diverse educational contexts.

## 5. References

Buitrago, M. & Chiappe, A., 2019. Representation of knowledge in digital educational environments: A systematic review of literature.. *Australasian Journal of Educational Technology,* 4(35), pp. 46-62.

Celik, B. & Cagiltay, K., 2024. Uncovering MOOC Completion: A Comparative Study of Completion Rates from Different Perspective. *Open Praxis,* 29 August, 16(3), p. 445–456.

Charntaweekhun , K. & Wangsiripitak, . S., 2006. Visual Programming using Flowchart. *2006 International Symposium on Communications and Information Technologies, Bangkok, Thailand,* pp. 1062-1065.

George, G. & La, . A. . M., 2024. PERKC: Personalized kNN With CPT for Course Recommendations in Higher Education. *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES,* Volume 17, pp. 885-890.

Islam, M. S. & Hosen , A. S. M. S., 2022. Personalized Course Recommendation System: A Multi-Model Machine Learning Framework for Academic Success. *Digital ,* 14(5), p. 2907.

Khan, M. A. Z. & Polyzou, A., 2024. Session-Based Course Recommendation Using Learner Interaction Sequences. *Electronics,* 3762(18), p. 13.

Liu, J., Zhang, H. & Liu, Z., 2020. Research on Online Learning Resource Recommendation Method Based on Wide & Deep and Elmo Model. *Journal of Physics: Conference Series,* 1437(1).

Peterson, R., 2013. *Why Do Students Drop Out of MOOCs?.* [Online] Available at: https://www.nas.org/blogs/article/why_do_students_drop_out_of_moocs
[Accessed 10 December 2025].

Ren, X. et al., 2022. A Deep Learning Framework for Multimodal Course Recommendation Based on LSTM+Attention. *Sustainability,* 14(12), p. 2907.

ScienceDirect, 2012. *State Transition Diagram.* [Online] Available at: https://www.sciencedirect.com/topics/computer-science/state-transition-diagram
[Accessed 13 December 2025].

Shah, D., 2021. *By The Numbers: MOOCs in 2021.* [Online] Available at: https://www.classcentral.com/report/mooc-stats-2021/
[Accessed 10 December 2025].

Tilahun, L. A. & Sekeroglu, B., 2020. An Intelligent Course Advising System Based on Expert Systems. *SN Computer Science,* 1(4), p. 1–15.

Xu, J., Xing, T. & Schaar, M. v. d., 2016. Personalized Course Sequence Recommendations. *IEEE Transactions on Signal Processing,* 64(20), p. 5340–5352.

Ziegler, N. et al., 2017. Interdisciplinary Research at the Intersection of CALL, NLP, and SLA: Methodological Implications From an Input Enhancement Project. *Journal of Learning Analytics,* 4(1), pp. 1-15.