

Machine Learning and Deep Learning Approaches for Recommendation Systems

Research Methodology in Computer Science (AI800)

Prepared by: **Ghada Alenazi**
Student ID: **261002011**

Submitted to:
Dr. Mohammed Anbar

Department of Computer Science
Master of Artificial Intelligence Program

MidoceanUniversity

April 2026

Table of Contents

Abstract.....	3
Chapter 1: Introduction.....	5
1.1 Introduction	5
1.2 Background	6
1.3 Problem Statement	7
1.4 Research Objectives	7
1.5 Research Scope	8
Chapter 2: Literature Review	9
2.1 Wide & Deep Learning for Recommendation Systems (Cheng et al., 2016).....	9
2.2 Neural Collaborative Filtering (He et al., 2017)	10
2.3 Hybrid CTR Prediction Model (Al-Hindy et al., 2019)	10
2.4 LightGCN (He et al., 2020)	11
2.5 Context-Aware Recommendation Systems (Al-Yafi et al., 2020).....	11
2.6 Self-Attentive Sequential Recommendation (Kang & McAuley, 2018).....	12
2.7 BERT4Rec Sequential Recommendation (Sun et al., 2019)	12
2.8 Deep Neural Networks for YouTube Recommendations (Covington et al., 2016)	12
2.9 Bayesian Personalized Ranking (Rendle et al., 2009)	13
2.10 Critical Analysis.....	13
2.11 Research Gap Identification.....	13
Chapter 3: Research Methodology	14
3.1 Research Design.....	14
3.2 Dataset Collection.....	15
3.3 Data Preprocessing	16
3.4 Model Development	17
3.5 Model Evaluation	19
3.6 Comparative Analysis	19
3.7 Tools and Technologies Used	20
3.8 Expected Outcomes	20
3.9 Gantt Chart	21
3.10 Ethical Considerations.....	21
3.11 Methodology Justification	22
References	23

Abstract

Recommendation systems have become a fundamental part of modern digital platforms such as e-commerce websites, streaming services, and personalized media applications. Their purpose is to help users discover relevant items by analyzing past behavior and interaction patterns. In recent years, machine learning and deep learning methods have significantly improved recommendation accuracy by modeling complex relationships between users and items. However, many advanced models require high computational resources, which makes them difficult to deploy in real-time environments or on devices with limited processing capacity.

Lightweight recommendation models improve efficiency, but they often fail to incorporate contextual information and changes in user preferences over time. This creates an important challenge in recommendation research: how to balance prediction accuracy, computational efficiency, contextual awareness, and temporal adaptability within a single framework.

This study examines major recommendation approaches, including Wide & Deep Learning, Neural Collaborative Filtering, Hybrid CTR models, LightGCN, context-aware recommendation systems, and sequential self-attention models. Based on the limitations identified in these approaches, this research proposes a lightweight recommendation framework that combines user–item

interactions, contextual features, and sequential behavioral patterns in a scalable architecture.

The proposed framework will be evaluated using benchmark datasets such as MovieLens and Amazon Product Review datasets. Performance will be measured using AUC, Precision@K, Recall@K, NDCG, training time, and inference latency. The expected outcome is a recommendation model that improves personalization quality while maintaining computational efficiency suitable for real-time applications.

Keywords: Recommendation Systems, Deep Learning, Machine Learning, Context-Aware Recommendation, LightGCN, Sequential Recommendation

Chapter 1: Introduction

1.1 Introduction

Recommendation systems are widely used in modern digital environments such as online marketplaces, streaming platforms, social media networks, and personalized content services. These systems assist users in finding products, services, or information that match their interests, making digital experiences more efficient and personalized.

As digital content continues to expand rapidly, recommendation systems have become increasingly important in helping users navigate large volumes of information. Traditional recommendation techniques such as collaborative filtering and matrix factorization provided early solutions, but their ability to model complex nonlinear interactions remained limited.

Recent advances in machine learning and deep learning have introduced more sophisticated recommendation models capable of learning deeper behavioral patterns. Models such as Wide & Deep Learning, Neural Collaborative Filtering, and LightGCN have shown strong improvements in recommendation performance. Despite these advances, many of these architectures require substantial computational power, limiting their use in real-time systems and resource-constrained devices.

This study focuses on addressing this challenge by exploring recommendation models that combine strong predictive performance with lightweight computational design.

1.2 Background

The development of recommendation systems has evolved significantly over the past two decades. Early systems mainly relied on collaborative filtering and content-based filtering methods. While these techniques were effective in smaller environments, they often struggled with sparse datasets and limited scalability.

The introduction of deep learning transformed recommendation research by enabling systems to automatically learn latent patterns from large-scale interaction data. Wide & Deep Learning combines memorization with generalization, while Neural Collaborative Filtering improves nonlinear interaction learning.

To improve efficiency, graph-based models such as LightGCN simplified recommendation architectures while maintaining strong predictive accuracy. At the same time, sequential recommendation models such as SASRec and BERT4Rec demonstrated the importance of modeling how user preferences change over time.

Although these advances improved recommendation quality, integrating efficiency, contextual awareness, and temporal adaptability into a single scalable framework remains a major unresolved challenge.

1.3 Problem Statement

Modern recommendation systems are expected to provide accurate and personalized suggestions while operating efficiently in real-time environments. Although advanced deep learning models achieve strong predictive performance, many require high computational resources, limiting their practical use in mobile devices and low-resource systems.

Lightweight models reduce computational complexity, but they often fail to incorporate contextual factors such as time, location, and device type. In addition, many efficient architectures cannot adequately model how user preferences evolve over time.

As a result, there is still no unified recommendation framework that successfully balances recommendation accuracy, computational efficiency, contextual understanding, and temporal adaptability. This study addresses that gap by proposing a lightweight recommendation framework that integrates these elements into a scalable design.

1.4 Research Objectives

This study aims to design an efficient recommendation framework that balances predictive quality and computational performance.

The specific objectives are:

1. To analyze existing recommendation models in terms of recommendation accuracy and efficiency.
2. To identify limitations in handling contextual and temporal information.
3. To design a framework integrating user–item interactions, contextual features, and sequential behavior.
4. To develop a lightweight model suitable for real-time recommendation environments.
5. To evaluate the proposed model using standard recommendation metrics.

1.5 Research Scope

This research focuses on machine learning and deep learning recommendation models related to:

- user–item interaction modeling,
- contextual recommendation systems,
- sequential behavioral learning.

The study compares major recommendation architectures using benchmark datasets such as MovieLens and Amazon Product Review datasets. Evaluation is limited to recommendation effectiveness and computational efficiency and does not extend to commercial-scale industrial deployment.

Chapter 2: Literature Review

Recommendation systems have developed rapidly in recent years due to advances in machine learning and deep learning. Researchers have introduced a wide range of models to improve recommendation accuracy, personalization, and scalability. Despite these improvements, balancing prediction quality with computational efficiency remains one of the main challenges in this field. This chapter reviews the most relevant studies related to this research and highlights how each contributes to understanding the current research gap.

2.1 Wide & Deep Learning for Recommendation Systems (Cheng et al., 2016)

Cheng et al. proposed the Wide & Deep Learning model to combine memorization and generalization within a unified recommendation framework. The wide component captures frequent feature interactions, while the deep component learns latent nonlinear patterns through embedding layers.

This hybrid structure improved recommendation performance in large-scale systems such as Google Play. Its main strength lies in combining linear and nonlinear learning effectively. However, the model requires high computational resources due to its deep architecture,

which limits efficiency in lightweight deployment environments.

This study is relevant because it demonstrates how recommendation accuracy can be improved while also revealing the computational burden that lightweight frameworks aim to reduce.

2.2 Neural Collaborative Filtering (He et al., 2017)

Neural Collaborative Filtering replaced traditional matrix factorization with neural architectures to better model nonlinear user–item relationships. By applying multilayer perceptrons, the model captures interaction patterns that conventional methods often miss.

NCF significantly improves recommendation precision, especially in sparse datasets. However, its deeper layers increase training complexity and processing time.

This model contributes to the present study by emphasizing the importance of nonlinear interaction modeling in recommendation systems.

2.3 Hybrid CTR Prediction Model (Al-Hindy et al., 2019)

Al-Hindy et al. introduced a hybrid deep learning architecture for click-through rate prediction that combines factorization machines with neural networks.

Although the model achieves strong predictive performance, it requires considerable hardware resources, which affects scalability in lightweight systems.

Its relevance lies in demonstrating how richer feature learning improves recommendation quality while increasing complexity.

2.4 LightGCN (He et al., 2020)

LightGCN simplified graph convolutional recommendation models by removing transformation layers and nonlinear activations.

This improves efficiency while maintaining strong performance. However, LightGCN does not fully integrate contextual or temporal information.

This limitation directly relates to the research gap addressed in this study.

2.5 Context-Aware Recommendation Systems (Al-Yafi et al., 2020)

Context-aware recommendation systems improve personalization by incorporating factors such as time, location, and device type.

These models increase relevance under changing conditions but often introduce additional computational complexity.

This study supports the current research objective of integrating contextual information without sacrificing efficiency.

2.6 Self-Attentive Sequential Recommendation (Kang & McAuley, 2018)

SASRec captures evolving user preferences using self-attention mechanisms.

It performs strongly in modeling behavioral sequences but requires higher memory resources.

This model is important because temporal adaptability is central to the current study.

2.7 BERT4Rec Sequential Recommendation (Sun et al., 2019)

BERT4Rec applies bidirectional transformer learning to recommendation systems.

It improves sequential prediction accuracy significantly, though transformer architectures remain computationally expensive.

This reinforces the need for lightweight alternatives.

2.8 Deep Neural Networks for YouTube Recommendations (Covington et al., 2016)

Covington et al. demonstrated how deep neural recommendation systems can scale for millions of users.

Although highly effective, such architectures require massive computational infrastructure.

This highlights the importance of designing scalable but efficient recommendation systems.

2.9 Bayesian Personalized Ranking (Rendle et al., 2009)

BPR introduced a ranking-based optimization method for implicit feedback recommendation tasks.

It is computationally efficient but lacks contextual and temporal modeling depth.

Its ranking strategy remains highly relevant for lightweight recommendation frameworks.

2.10 Critical Analysis

The reviewed studies show a clear trade-off: highly accurate deep learning models often increase computational cost, while lightweight models improve efficiency but sacrifice contextual and temporal depth.

No existing model fully combines:

- high recommendation accuracy,
- low computational complexity,
- contextual awareness,
- temporal adaptability.

This unresolved limitation defines the main research gap addressed in this study.

2.11 Research Gap Identification

Current recommendation systems still face four major limitations:

1. High computational cost in advanced deep learning models.

2. Weak integration of contextual variables in lightweight architectures.
3. Limited ability to model changing user preferences over time efficiently.
4. Lack of unified frameworks balancing efficiency and personalization quality.

Therefore, this research proposes a lightweight recommendation framework that combines these essential components in a scalable architecture.

Chapter 3: Research Methodology

This chapter explains the methodology used to achieve the objectives of this study. Since the main purpose of the research is to design a lightweight, context-aware, and temporally adaptive recommendation framework, a comparative experimental research design is adopted. This approach allows systematic evaluation of the proposed model against existing recommendation systems.

3.1 Research Design

This study follows a comparative experimental design in which the proposed recommendation framework is compared with baseline recommendation models in terms of:

- recommendation accuracy,
- computational efficiency,
- contextual awareness,
- temporal adaptability.

The research process is divided into five main phases:

1. Dataset Collection
2. Data Preprocessing
3. Model Development
4. Model Evaluation
5. Comparative Analysis

This structure ensures a consistent and measurable evaluation process.

3.2 Dataset Collection

To ensure reliable benchmarking, this study uses widely recognized public datasets commonly applied in recommendation system research.

Selected Datasets:

- MovieLens Dataset
- Amazon Product Review Dataset

These datasets were selected because they:

- contain large-scale user–item interaction records,
- support benchmark comparison with previous studies,
- include rich recommendation-related metadata.

Each dataset includes:

- user identifiers,
- item identifiers,
- ratings or interaction values,
- timestamps,
- contextual attributes where available.

These datasets provide sufficient diversity for evaluating both recommendation accuracy and scalability.

3.3 Data Preprocessing

Before model training, raw data must be cleaned and transformed into suitable input format.

The preprocessing stage includes:

1. Removing duplicate and incomplete records.
2. Normalizing rating values.
3. Encoding categorical variables numerically.
4. Extracting contextual attributes such as:
 - time,
 - location (if available),
 - device type.
5. Generating sequential interaction histories for each user.

This preprocessing improves data consistency and enhances model learning quality.

3.4 Model Development

This phase involves implementing both baseline models and the proposed recommendation framework.

Baseline Models:

- Wide & Deep Learning
- Neural Collaborative Filtering (NCF)
- Hybrid CTR Model
- LightGCN
- Context-Aware Recommendation Model
- SASRec
- BERT4Rec

Proposed Model Framework:

The proposed framework integrates four main components:

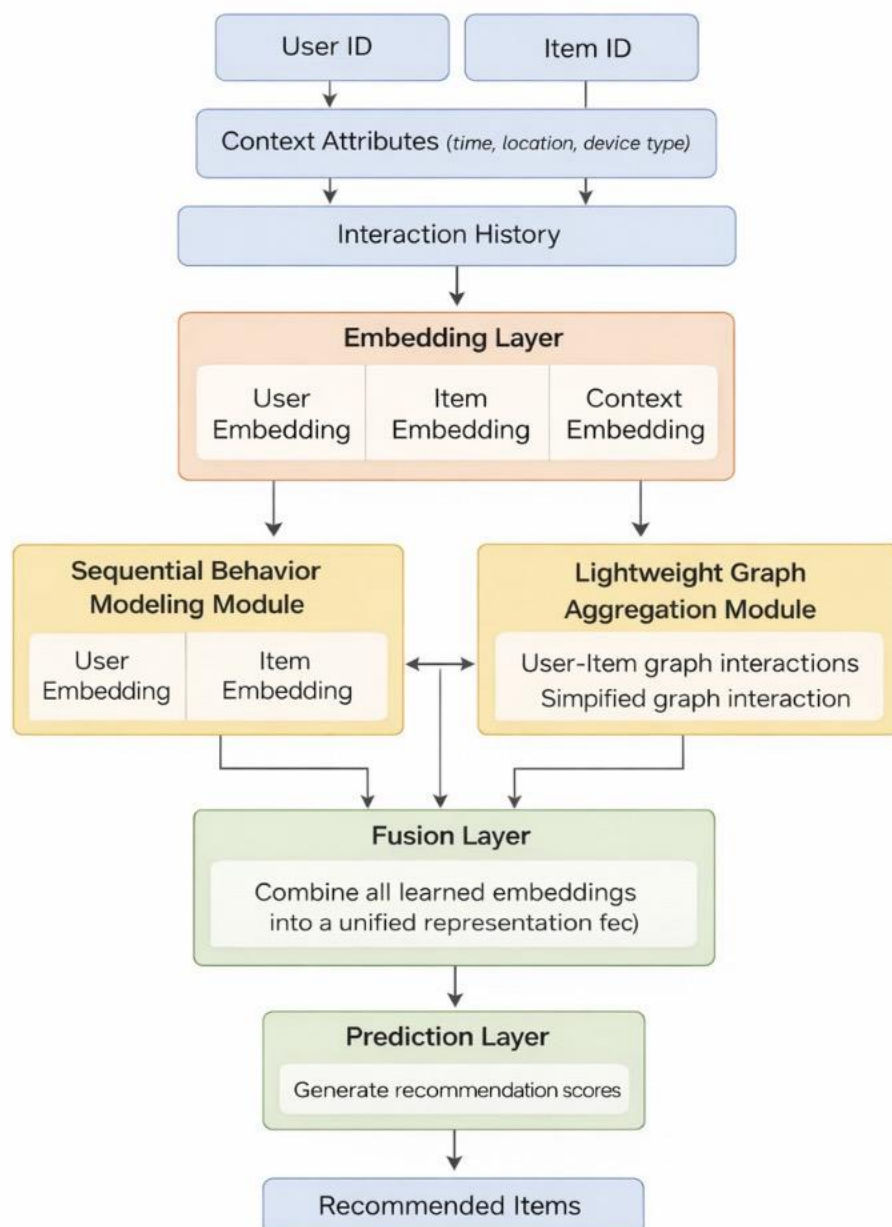
1. User–item interaction embeddings
2. Contextual feature embeddings
3. Sequential behavioral modeling
4. Lightweight graph aggregation mechanism

The goal is to combine:

- strong predictive accuracy
- reduced computational cost

- contextual relevance
- temporal adaptability

This design directly addresses the research gap identified in Chapter 2.



The model processes user–item interactions, contextual attributes, and interaction history as inputs. These are

converted into embeddings, analyzed through sequential and graph modules, then fused and passed to the prediction layer to generate recommendation scores and final recommended items.

3.5 Model Evaluation

Each model will be evaluated using standard recommendation metrics to measure both performance quality and efficiency.

Performance Metrics:

- AUC (Area Under Curve)
- Precision@K
- Recall@K
- NDCG@K
- Hit Ratio (HR)

Efficiency Metrics:

- Training time
- Inference latency
- Memory usage

These metrics provide balanced assessment between predictive strength and system efficiency.

3.6 Comparative Analysis

After evaluation, the proposed framework will be compared against baseline models.

The comparison focuses on:

1. Recommendation accuracy improvement
2. Reduction in computational complexity

3. Context integration effectiveness
4. Temporal adaptability performance

This phase determines whether the proposed model achieves a better balance than existing approaches.

3.7 Tools and Technologies Used

The implementation environment includes:

- Python Programming Language
- TensorFlow / PyTorch
- Pandas
- NumPy
- Scikit-learn
- Jupyter Notebook

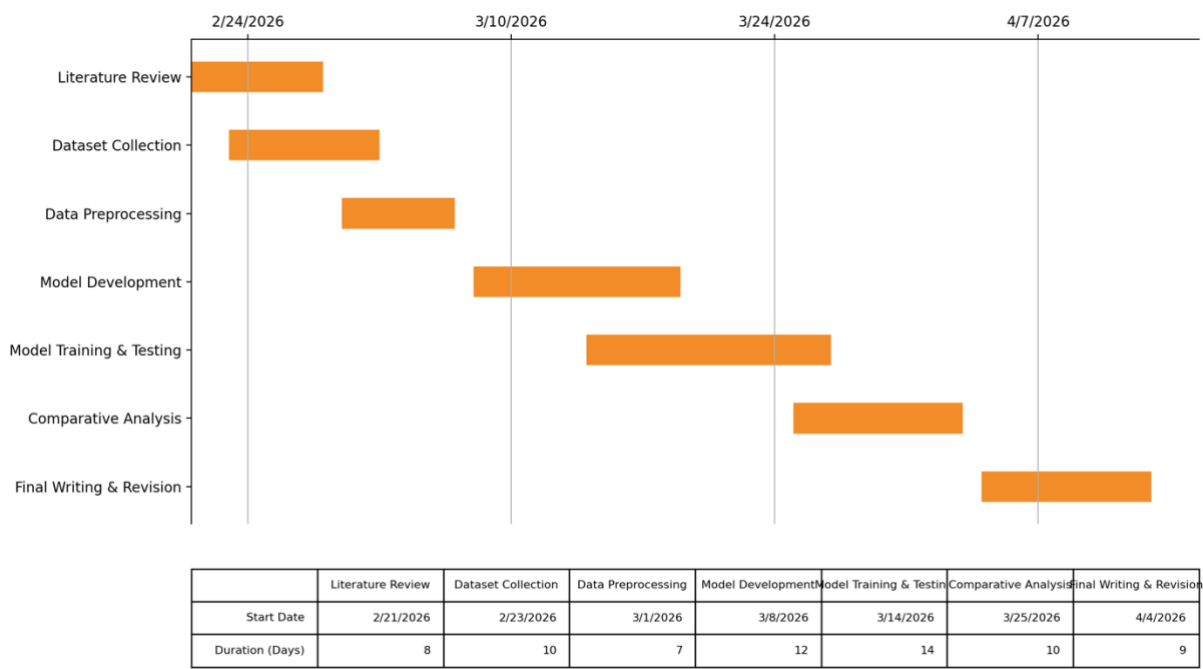
These tools provide flexibility for machine learning experimentation and reproducibility.

3.8 Expected Outcomes

This study is expected to achieve the following outcomes:

1. Development of a lightweight recommendation framework.
2. Improved balance between efficiency and recommendation accuracy.
3. Better contextual personalization quality.
4. Stronger modeling of dynamic user preferences.
5. Lower computational cost than heavy deep learning architectures.

3.9 Gantt Chart



3.10 Ethical Considerations

This research uses publicly available benchmark datasets that do not expose private personal identities.

Ethical standards are maintained by:

- respecting dataset licenses,
- protecting anonymity of user data,
- avoiding misuse of sensitive information.

Ethical compliance ensures responsible research practice.

3.11 Methodology Justification

This methodology is appropriate because it:

- directly supports the research objectives,
- enables measurable model comparison,
- aligns with accepted standards in recommendation system research.

The chosen approach provides reliable evidence for evaluating the effectiveness of the proposed framework.

References

- Al-Hindy, A., Al-Sayid, B., & Al-Ani, A. (2019). A hybrid deep learning architecture for click-through rate prediction. *IEEE Access*.
- Al-Yafi, H., Galland, A., & Gaussier, E. (2020). Deep learning-based context-aware recommendation system. In *Advances in Information Retrieval* (pp. 317–330). Springer.
- Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Fu, W., & Shah, H. (2016). Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems* (pp. 7–10). ACM.
- Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191–198).
- He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (2020). LightGCN: Simplifying and powering graph convolution network for recommendation. In *Proceedings of SIGIR 2020* (pp. 639–648).
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). Neural collaborative filtering. In *Proceedings of WWW 2017* (pp. 173–182).
- Kang, W.-C., & McAuley, J. (2018). Self-attentive sequential recommendation. In *Proceedings of ICDM 2018*.
- Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2009). BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of UAI 2009*.
- Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. (2019). BERT4Rec: Sequential recommendation with

bidirectional encoder representations from transformer.
In Proceedings of CIKM 2019.