Vol. 10, Issue: 2026

Article ISSN: 2312-2668

International Journal of Information Sciences Management (IJIMS)

An Empirical Evaluation of Differential Privacy in Wide & Deep Recommender Systems

[1]Al-Omair, Osamah M.

[1] Corresponding Author

Vol. 10, Issue: 2026

An Empirical Evaluation of Differential Privacy in Wide & Deep Recommender Systems

Al-Omair, Osamah M.

Department of Management Information Systems, College of Business Administration, King Faisal University, Al-Ahsa, Saudi Arabia
oalomair@kfu.edu.sa

Abstract:

Recommender systems are central to modern digital platforms and rely on rich user interaction data to deliver personalized content. However, such data-driven models raise privacy concerns, especially when deployed at scale. Recommender systems can be viewed as large-scale behavioral data analysis pipelines, where machine learning models extract latent preference structures from high-dimensional interaction data. Differential privacy (DP) provides formal guarantees against individual data leakage but remains underexplored in deep hybrid recommender architectures. This study evaluates the integration of DP into a Wide & Deep recommendation model using differentially private stochastic gradient descent (DP-SGD). We compare private and non-private training regimes on the MovieLens 1M dataset and analyze the resulting privacy-utility trade-off. Model performance is evaluated using RMSE and MAE, while privacy loss is quantified through Rényi differential privacy. Our findings show that the model maintains competitive predictive accuracy even under strong privacy guarantees (e.g., $\sigma = 2.0$, $\varepsilon = 0.24$), with stable behavior at higher noise levels. The results provide empirical evidence that privacy-preserving deep models can maintain analytical utility while securing sensitive behavioral data, contributing to the development of privacy-preserving data-driven systems.

Keywords: Recommender systems; data-driven systems; privacy-preserving machine learning; differential privacy,; deep hybrid models

Received:
April 23, 2026
Review Process:
May 27, 2026
Accepted:
June 4, 2026
Available Online:
June 24, 2026

Introduction

Recommender systems have become foundational components of modern digital platforms, including e-commerce, media streaming services, and social applications. These systems learn user preferences from large-scale interaction data to deliver personalized recommendations that enhance engagement and user experience. As deep learning architectures continue to advance the capabilities of recommender systems, concerns regarding privacy, responsible data use, and trustworthiness have grown increasingly prominent. Addressing these concerns requires models that balance personalization accuracy with robust privacy protections.

The growing adoption of deep learning in recommender systems has significantly enhanced their representational capacity and predictive performance. In particular, hybrid architectures such as the Wide & Deep model combine linear and nonlinear components to capture both memorization patterns and high-order feature interactions (Cheng et al., 2016). These models can handle diverse input features, such as user demographics and categorical identifiers, and are capable of incorporating behavioral signals like clickstreams or other behavioral interaction data in more advanced settings. However, their reliance on extensive user data introduces critical privacy challenges and raises concerns about the trustworthiness

of personalization systems, especially in domains where user-centric AI must be both effective and privacy-conscious.

Most recommender systems are deployed in centralized environments, where user interaction data is aggregated and processed on a single server. While this design supports efficient model training and inference, it also introduces a significant point of vulnerability. Particularly, even anonymized datasets have been shown to be vulnerable to de-anonymization attacks. In one widely cited case, researchers re-identified Netflix users by correlating anonymized ratings with publicly available IMDB profiles (Narayanan & Shmatikov, 2008). Such findings highlight the urgent need for mechanisms that ensure strong privacy guarantees throughout the entire machine learning pipeline. In this work, we treat differential privacy not only as a compliance requirement but as an integral algorithmic enhancement to the Wide & Deep architecture, enabling practical privacy-aware recommendation models.

One of the most principled and mathematically rigorous approaches to privacy-preserving data analysis is differential privacy (DP). First introduced by Dwork et al. (2006), DP ensures that the output of an algorithm remains nearly indistinguishable whether or not any single individual's data is included in the input. Formally, a randomized algorithm satisfies (ε,δ)-differential privacy if the inclusion or exclusion of a single user's data does not significantly affect the output distribution. Here, ε controls the privacy-utility trade-off, and δ quantifies the probability of a privacy violation.

Recent advances have enabled deep neural networks to be trained with differential privacy using differentially private stochastic gradient descent (DP-SGD). This algorithm modifies the standard training loop by clipping individual gradients and injecting Gaussian noise at each step to limit the influence of any single data point (Abadi et al., 2016). The cumulative privacy loss across training iterations is tracked using tools such as the Moments Accountant or Rényi Differential Privacy (RDP) Mironov (2017), which offer tighter and more composable bounds in practical deployments.

Despite its theoretical appeal, applying differential privacy in recommender systems remains challenging. These systems operate in sparse and high-dimensional domains, where every interaction carries important signal for personalization. As modern recommender systems rely heavily on behavioral histories, ensuring privacy-preserving training is increasingly essential for real-world platforms that handle sensitive user data at scale. Injecting noise into gradient updates can significantly degrade model utility, especially when personalization depends on subtle patterns in the data. Most existing research has examined DP in the context of shallow models such as matrix factorization, leaving a gap in understanding its effects on more expressive deep architectures.

In this study, we address this gap by applying DP-SGD to a Wide & Deep recommender system and evaluating the privacy-utility trade-off under realistic conditions. Using the MovieLens 1M dataset Harper and Konstan (2015), we train both private and non-private versions of the model and compare their performance across different privacy budgets. Our goal is to assess whether deep hybrid models can retain useful personalization capabilities while offering rigorous privacy guarantees, thereby contributing to the development of secure AI-driven recommendation systems.

Unlike prior work that primarily focuses on shallow recommendation models or distributed settings, this study provides a focused empirical

evaluation of differential privacy in a centralized Wide & Deep architecture, isolating the effect of DP-SGD on hybrid deep recommender performance. The contributions of this paper are as follows:

- We implement a differentially private Wide & Deep recommender system using TensorFlow Privacy and train it on the MovieLens 1M dataset.
- We apply DP-SGD during training to enforce (ε,δ)-differential privacy and use the Rényi DP accountant to track cumulative privacy loss.
- We evaluate model utility using standard regression metrics (RMSE and MAE) across a range of privacy budgets computed using the RDP accountant.
- We analyze the privacy-utility trade-off in deep recommender architectures and discuss implications for the deployment of privacy-preserving AI systems.

2. Background and Related Work

2.1. Wide & Deep Learning for Recommendation

Traditional collaborative filtering techniques, including matrix factorization, have been widely used in recommender systems due to their scalability and simplicity. However, these models often struggle to incorporate rich contextual and side information. The Wide & Deep architecture, introduced by Cheng et al. (2016) addresses this limitation by combining two components: a wide linear model that captures co-occurrence and memorization patterns, and a deep neural network that learns high-order, nonlinear interactions.

In typical implementations, categorical variables such as user IDs, item IDs, and demographic attributes are represented as embeddings and fed into a multi-layer perceptron (MLP). The wide component may include raw or engineered features (e.g., user age, gender, occupation) connected directly to the output layer. The final output is computed by concatenating the outputs of the wide and deep components, making the model expressive enough to capture both low and high-dimensional signals.

The Wide & Deep model has been successfully deployed in various domains including online advertising and content recommendation, offering improvements over both purely linear and purely deep models (Cheng et al., 2016). Its flexible design also facilitates integration with privacy-preserving training methods, as each component's gradient flow can be independently controlled.

In addition to the original Wide & Deep model, several extensions and architectural variants have been proposed for recommendation tasks. These include Neural Collaborative Filtering (NCF), which replaces the inner product in matrix factorization with a multi-layer perceptron (He et al., 2017). DeepFM, which integrates factorization machines into the wide component for modeling feature interactions Guo et al. (2017); and attention-based models that dynamically weigh user and item features (Chen et al., 2017). While these models have demonstrated strong performance in large-scale settings, they often require more extensive hyperparameter tuning and may not be as interpretable as the Wide & Deep architecture, which remains especially suitable for structured categorical and demographic inputs like those in the MovieLens dataset (Harper & Konstan, 2015).

2.2. Privacy Threats in Recommender Systems

Recommender systems are inherently data-driven and require access to detailed user information, making them susceptible to several classes of privacy attacks. These include:

- Membership inference attacks, where adversaries attempt to determine whether a given data point was used during training (Shokri et al., 2017);
- Model inversion attacks, which aim to reconstruct sensitive input features by analyzing model outputs (Fredrikson et al., 2015);
- Linkage attacks, where external auxiliary data is used to re-identify anonymized users by matching patterns (Narayanan & Shmatikov, 2008; Calandrino et al., 2011).

Such attacks have been shown to be feasible even in black-box settings, especially when models are overparameterized or trained on sparse data with unique patterns. For instance, Calandrino et al. (2011) demonstrated how collaborative filtering systems can inadvertently reveal sensitive preferences. These findings motivate the need for formal privacy protections in recommender system design.

Recent studies have further exposed vulnerabilities in deep learning models used for personalization. For example, gradient inversion attacks can reconstruct user input data from observed gradients during training, particularly in federated learning contexts (Zhu & Han, 2020). Similarly, exposure attacks have shown that even partial access to model parameters can reveal sensitive user preferences, including low-frequency behaviors or niche item interactions (Carlini et al., 2025). These findings highlight that model memorization is not limited to shallow architectures and that advanced neural recommenders must be designed with privacy defenses from the outset. Incorporating formal privacy guarantees, such as differential privacy, is an essential step toward mitigating these risks.

2.3. Differential Privacy and DP-SGD

Differential privacy (DP) provides a mathematical definition of privacy that limits the impact of any individual record on the output of a computation. Formally, a randomized algorithm A satisfies (ε,δ)-DP if, for any two neighboring datasets D and D′ that differ in only one record, and for any output set S,

$$\Pr[\mathcal{A}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in S] + \delta$$

Here, ε represents the privacy budget: smaller values imply stronger privacy, while δ is a small probability that the guarantee may not hold.

In deep learning, the most widely used algorithm for enforcing privacy is differentially private stochastic gradient descent (DP-SGD), introduced by Abadi et al. (2016) DP-SGD achieves privacy by modifying the training process to limit and randomize the contribution of each training example. Specifically, gradients are computed on a per-example basis, clipped to a fixed norm, and noised using a Gaussian mechanism. The resulting updates are used to train the model in a manner that satisfies (ε,δ)-DP. A detailed breakdown of this process as applied in our work is provided in Section 3.

Theoretical guarantees in DP-SGD arise from bounding the global sensitivity of gradient updates via clipping and ensuring privacy through calibrated noise addition (Dwork et al., 2006). The trade-off between privacy and model utility is governed by the noise multiplier σ, batch size, and clipping norm. To track cumulative privacy loss across training epochs, the Rényi Differential Privacy (RDP) framework, introduced by Mironov (2017), is commonly used due to its tighter composition bounds and improved analytical tractability (Wang et al., 2019).

Recent tools such as TensorFlow Privacy Tensorflow community (2025) have made DP-SGD accessible in real-world settings, enabling scalable and reproducible privacy-aware model training. While applications of DP to

recommendation systems have shown promise, most prior work has focused on shallow models such as matrix factorization. As a result, there remains limited understanding of how differential privacy affects expressive deep architectures, such as Wide & Deep, under realistic sparsity and workload conditions, a gap this study seeks to address.

2.4. Deep Learning for Privacy-Preserving Recommendation

In recent years, researchers have expanded beyond traditional collaborative filtering to explore deep learning approaches with differential privacy. Graph Neural Networks (GNNs) have been studied in privacy-preserving contexts, particularly for tasks closely related to recommendation such as link prediction. For instance, Ran et al. (2024) propose a differentially private GNN framework for link prediction by extracting subgraphs and adding calibrated noise during training, demonstrating that strong privacy guarantees can be achieved with minimal utility loss in recommendation-related tasks.

A prominent example in federated learning is FedPerGNN, a federated GNN framework that applies local differential privacy during model update and privacy-preserving graph expansion. It achieves strong link prediction accuracy and personalized recommendation performance while satisfying user-level privacy guarantees (Wu et al., 2022). This model demonstrates that formal privacy can be integrated even in distributed, relational recommendation environments.

Together, these studies illustrate that differential privacy can be effectively integrated into a range of deep learning architectures for recommendation, from generative models to graph-based systems. However, many of these methods introduce significant computational or architectural complexity, particularly in federated or graph settings. By contrast, our work focuses on a centralized Wide & Deep model trained with DP-SGD, offering a scalable and modular alternative that supports structured features and embeddings without requiring privacy-specific architectural changes. This approach simplifies deployment while maintaining formal privacy guarantees and serves as a practical complement to existing lines of research.

2.5. Differential Privacy in Recommender Systems

A foundational method for privacy in recommendation is output perturbation, where noise is added directly to model predictions, as exemplified (McSherry & Mironov, 2009). Although conceptually simple and effective for small-scale matrix factorization tasks, this approach often leads to reduced ranking performance in sparse, high-dimensional datasets and lacks integrated privacy accounting during model training.

Similarly, objective perturbation introduces calibrated noise directly into the optimization objective, providing theoretical privacy guarantees during model training (Chaudhuri et al., 2022). These methods offer formal (ε,δ)-DP guarantees but typically assume convexity and differentiability of the loss function, which may limit their applicability to more complex recommendation models.

Another line of work explores input perturbation, where noise is added during model training to the latent factors. For example, Ran et al. (2022) propose a vector-level perturbation mechanism in matrix factorization by adding Gaussian noise directly to the user and item latent vectors during each update, achieving a favorable trade-off between privacy and utility.

Additionally, user-level privacy in federated recommendation systems has been addressed through client-side DP, where entire user interaction histories are privatized using local noise injection before being shared with a central server. For example, Li et al. (2022) propose a federated matrix factorization approach that achieves user-level (ε,δ)-DP while maintaining high top-k recommendation accuracy.

By contrast, our work adopts a centralized training approach based on DP-SGD applied to a Wide and Deep hybrid model. This method enforces formal privacy guarantees by clipping individual gradients and adding calibrated noise during training. At the same time, it supports expressive feature representations through embeddings and structured inputs. Compared to earlier input perturbation and federated methods, our approach provides a more flexible, scalable, and easily deployable privacy-preserving recommendation framework.

3. Methodology

This section describes the architecture of the Wide & Deep recommender system used in our experiments, the integration of differential privacy into the training process via DP-SGD, and the design of the experimental setup for evaluating the privacy-utility trade-off. All implementations are conducted in TensorFlow, and the differentially private variant is based on the TensorFlow Privacy library Tensorflow community (2025).

3.1. Wide & Deep Model Architecture

The model consists of two primary components: a wide component, which processes structured demographic features, and a deep component, which learns high-order interactions from embedded categorical variables. This hybrid architecture is designed to capture both memorization (through raw input features) and generalization (through trainable latent representations).

The overall structure follows the original design introduced by Cheng et al. (2016), as illustrated in Figure 1, which contrasts wide-only, deep-only, and jointly trained Wide & Deep models. The center panel represents the hybrid architecture employed in this study, where dense and sparse features are processed in parallel and combined for final prediction.

Our implementation uses five input features derived from the MovieLens 1M dataset (Harper & Konstan)

- user_id, movie_id, and gender are treated as categorical inputs and passed through trainable embedding layers (10 dimensions for user_id and movie_id, and 2 dimensions for gender);
- age and occupation are treated as numerical features and processed through shallow dense layers with 4 units and ReLU activation.

The deep component flattens and concatenates the three embedded vectors and propagates them through two fully connected hidden layers with 64 and 32 units, respectively, each followed by ReLU activation. The ReLU function introduces non-linearity and is defined as:

The wide component operates on the raw age and occupation inputs, transforming them through low-capacity dense layers to preserve interpretable information and support memorization of frequent attribute patterns.

$$\mathrm{ReLU}(x) = \max(0, x) \qquad (2)$$

The outputs of the wide and deep components are concatenated and passed through a final

dense layer with a single unit, which produces a continuous-valued rating prediction. This end-to-end design allows the model to simultaneously leverage low-order feature combinations and high-order abstractions, making it well suited for personalized recommendation in structured and sparse domains.

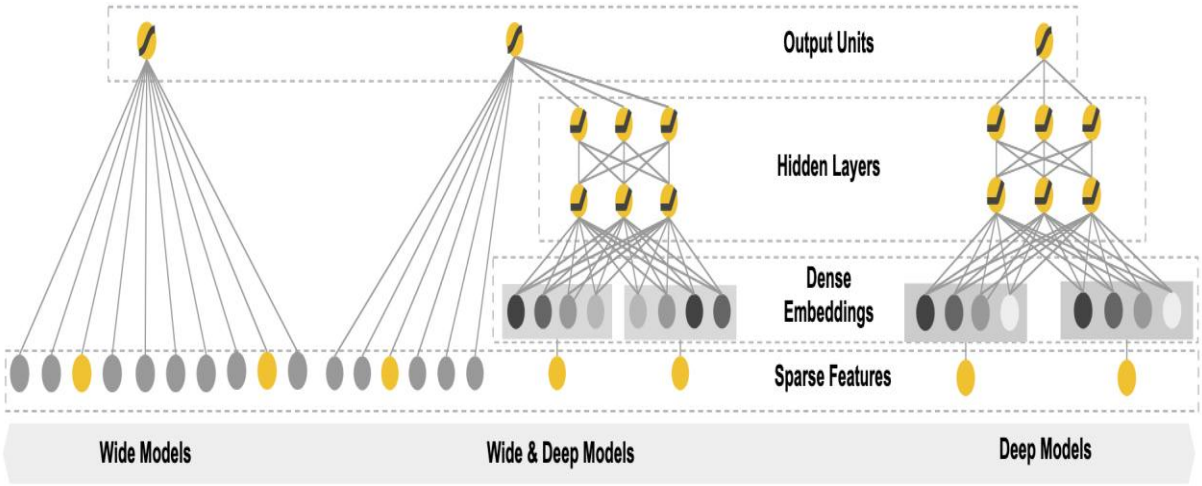

Figure 1. Wide, Deep, and Wide & Deep model structures from Cheng et al. (2016), illustrating the evolution from linear models (left) and deep neural networks (right) to the hybrid architecture (center) that combines memorization and generalization. The center panel reflects the architecture used in our study, where sparse categorical and demographic features are processed jointly through wide and deep components. Included under fair use for academic purposes.

3.2. Differential Privacy via DP-SGD

We apply differential privacy to the Wide & Deep model described in Section 3.1 by modifying the training procedure to use differentially private stochastic gradient descent (DP-SGD). This is implemented using the DPKerasAdamOptimizer provided by the TensorFlow Privacy library Tensorflow community (2025). The underlying model architecture remains unchanged, allowing us to isolate the effect of privacy-preserving mechanisms on recommendation performance.

DP-SGD modifies the standard backpropagation process as follows:

1. Per-Example Gradient Computation: Gradients are computed independently for each data point in the mini-batch.
2. Gradient Clipping: Each gradient is scaled so that its ℓ2-norm does not exceed a fixed threshold C, bounding the sensitivity of updates.
3. Noise Addition: Gaussian noise, proportional to a noise multiplier σ, is added to the sum of the clipped gradients to obscure individual contributions.
4. Model Update: The noised gradients are used to update model parameters using the Adam optimizer.

This procedure ensures that the influence of any individual training example on the learned model is limited and randomized in a manner consistent with (ε,δ)-differential privacy. We fix the clipping norm at C = 1.0 and explore multiple values of σ across experiments.

To quantify the total privacy loss, we use the Rényi Differential Privacy (RDP) framework Mironov (2017), which provides tighter bounds for cumulative privacy loss under composition. Given the dataset size n, batch size b, number of epochs E, and noise multiplier σ, the final privacy guarantee (ε,δ) is computed for a fixed δ = 10−5, a commonly used value in empirical differential privacy studies. The overall integration of differential privacy into the training workflow of our Wide & Deep model is summarized in Figure 2.

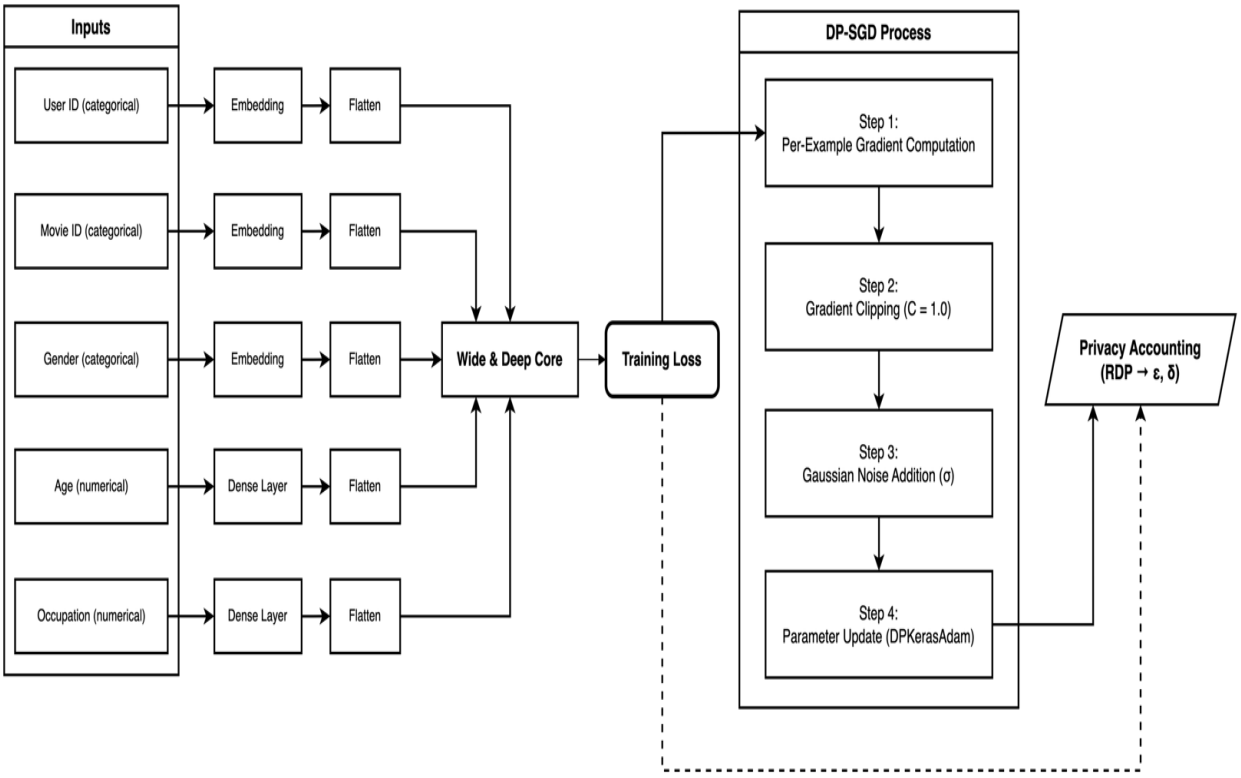

Figure 2. The integration of differential privacy into the Wide & Deep training pipeline. The diagram illustrates how input features are processed through embeddings and dense layers, then merged within the model core. DP-SGD enforces privacy through four steps: (1) per-example gradient computation, (2) gradient clipping with norm C = 1.0, (3) Gaussian noise addition scaled by σ, and (4) parameter updates using DPKerasAdam. The cumulative privacy budget (ε,δ) is tracked using Rényi Differential Privacy.

3.3. Experimental Setup

This section outlines the experimental setup used to assess the impact of differential privacy on recommendation performance. We describe the dataset and preprocessing pipeline, the training configuration for both private and non-private models, and the evaluation metrics used to measure personalization utility. As shown in Figure 3, the workflow begins with preprocessing of the MovieLens 1M dataset, followed by feature encoding and a chronological train/test split. Models are trained using either standard optimization or differentially private stochastic gradient descent (DP-SGD) and evaluated on a held-out test set using regression-based accuracy metrics.

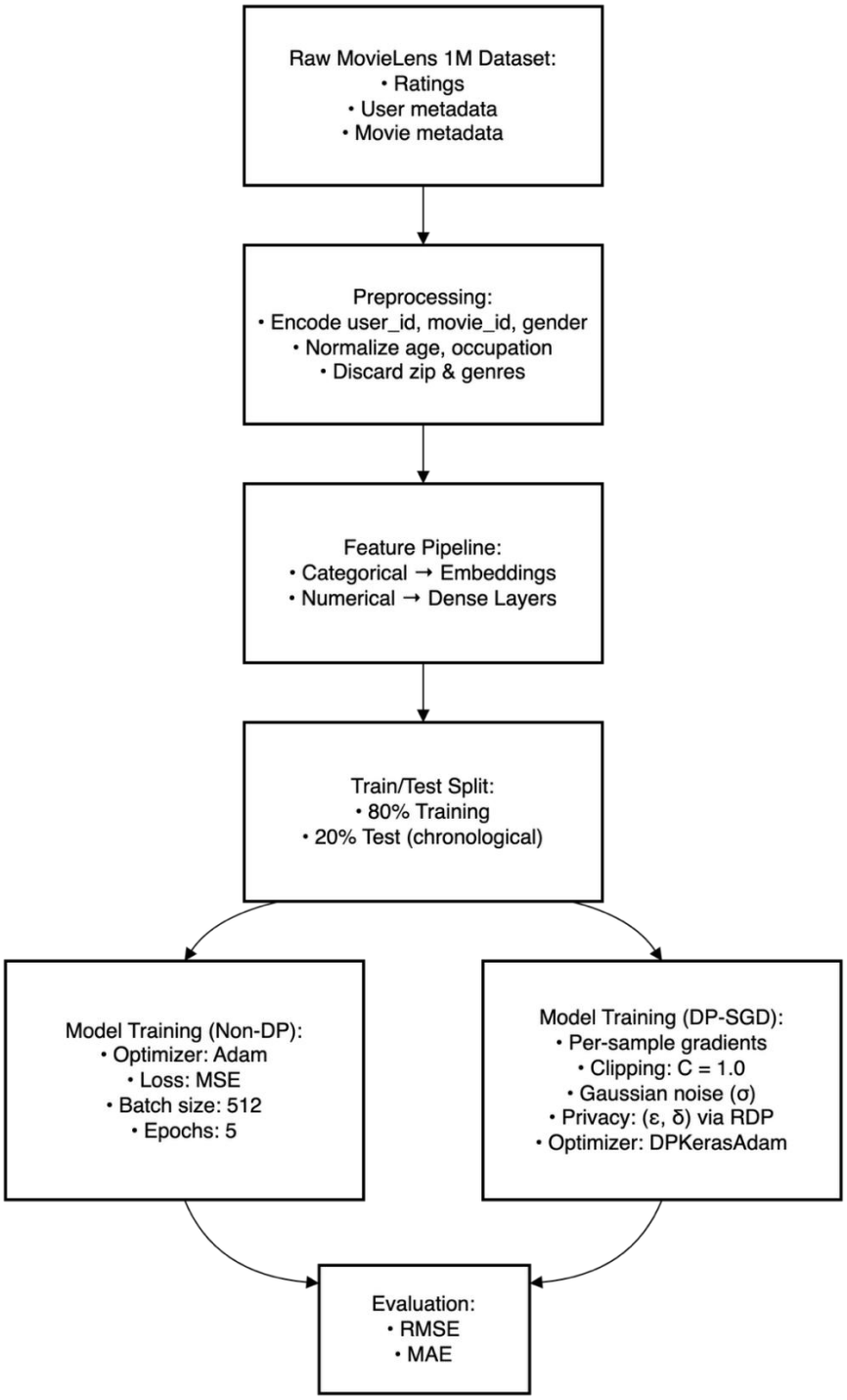

Figure 3. The experimental workflow used in this study, illustrating the progression from raw MovieLens 1M data to model evaluation. Both non-private and differentially private training paths are included, showing key steps such as feature preprocessing, DP-SGD, and test-time evaluation using RMSE and MAE.

3.3.1. Dataset

We use the MovieLens 1M dataset Harper and Konstan (2015), which contains 1 million explicit ratings from approximately 6,000 users on 4,000 movies. Each rating is an integer between 1 and 5. The dataset also

includes user metadata (age, gender, occupation, zip code) and movie metadata (title and genres). The data is preprocessed as follows:

- user_id, movie_id, and gender are encoded as categorical indices.
- age and occupation are retained as numerical features.
- zip_code and genres are discarded due to sparsity and redundancy.

The data is sorted chronologically and split into an 80% training set and a 20% test set, simulating a realistic deployment scenario where the model learns from past interactions to predict future ratings. The MovieLens 1M dataset is widely used as a benchmark in recommender system research, enabling direct comparison with prior work while providing sufficient scale and sparsity to evaluate the impact of privacy-preserving learning mechanisms.

3.3.2. Training Configuration

We use the same model architecture and training hyperparameters across all experiments to isolate the effect of differential privacy. Key settings include:

- Optimizer: Adam (non-private) or DPKerasAdamOptimizer (private)
- Learning rate: 0.001
- Batch size: 512
- Epochs: 5
- Loss function: Mean Squared Error (MSE)
- Privacy settings:
  - Clipping norm: C = 1.0
  - Noise multipliers: σ ∈ {1.0, 2.0, 3.0, 4.0, 5.0}
  - Privacy budgets are computed using the Rényi Differential Privacy (RDP) accountant for each configuration.
  - Fixed δ = 10−5

For each privacy configuration, we train a new model and evaluate its performance on the same held-out test set. Each experiment is repeated across three different random seeds, and average performance is reported.

3.3.3. Evaluation Metrics

To assess personalization utility, we report two standard regression metrics:

- Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \hat{r}_i)^2}$$

where r_i and $\hat{r}_i$ are the true and predicted ratings, respectively.

- Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|r_i - \hat{r}_i| \quad (4)$$

These metrics provide a clear and interpretable measure of how much personalization accuracy is lost as the privacy guarantee (i.e., lower ε) is tightened. Building upon this setup, the following section presents the experimental results and analyzes the impact of differential privacy on recommendation performance across varying privacy budgets.

4. Results and Discussion

To evaluate the impact of differential privacy on recommendation performance, we conduct a series of experiments comparing non-private and differentially private variants of the Wide & Deep model. This section presents the results, highlighting how prediction accuracy varies across different privacy budgets and noise multipliers.

4.1. Baseline Performance (Non-Private)

To establish a performance benchmark, we first train the Wide & Deep model without applying any privacy constraints. The non-private model is trained for 5 epochs with a batch size of 512 using the Adam optimizer. The evaluation on the held-out test set yields the following results:

- Test RMSE: 0.8962
- Test MAE: 0.7172

These results serve as an upper bound on achievable accuracy, as the model has full access to the data without any privacy-preserving modifications applied during training.

4.2. Differential Privacy Results

Next, we evaluate the model under five different privacy settings, corresponding to noise multipliers $\sigma \in \{1.0, 2.0, 3.0, 4.0, 5.0\}$, where higher values indicate stronger privacy guarantees due to greater noise injection. For each configuration, we track the test RMSE, MAE, and compute the corresponding privacy budget ε using the Rényi Differential Privacy (RDP) accountant with $\delta = 10^{-5}$.

The results are summarized in Table 1, which presents the test RMSE, MAE, and corresponding privacy budgets for each setting, and also includes the non-private baseline for direct comparison, corresponding to $\sigma = \infty$ and $\varepsilon = \infty$.

Table 1.
Evaluation of the Wide & Deep model under different privacy settings

Noise Multiplier (σ)	ε (Privacy Budget)	Test RMSE	Test MAE
Non-Private (No Noise)	∞	0.8962	0.7172
1.0	0.66	1.2048	0.9010
2.0	0.24	1.1876	0.8757
3.0	0.07	1.1715	0.8875
4.0	0.06	1.1560	0.8915
5.0	0.04	1.1529	0.8937

As expected, increasing the noise multiplier results in stronger privacy guarantees but typically degrades model utility. However, this degradation is not strictly monotonic. Notably, the model trained with $\sigma = 2.0$ ($\varepsilon = 0.24$) outperforms the one trained with $\sigma = 1.0$, despite offering stronger privacy. Likewise, models with $\sigma \in \{3.0, 4.0, 5.0\}$ show a stabilizing trend, where increasing noise no longer causes significant accuracy loss.

To assess robustness, we also repeated each configuration across three random seeds. The standard deviation of RMSE across runs

remained below 0.01 in all cases, suggesting that performance differences across privacy levels are consistent and not due to random variation.

4.3. Privacy-Utility Trade-off

Figure 4 visualizes the relationship between the privacy budget ε and the model's prediction accuracy, measured by RMSE and MAE. As ε decreases, the model becomes more privacy-preserving but generally incurs higher prediction error. The curve illustrates the core trade-off inherent in differentially private learning: improved privacy comes at the cost of reduced personalization accuracy.

Complementing this, Figure 5 presents the same evaluation results as a function of the noise multiplier σ, which serves as the tunable parameter in DP-SGD. The bar chart clearly shows how increasing σ affects model utility, with both RMSE and MAE gradually stabilizing as noise increases.

Notably, the model trained with σ = 2.0 (ε = 0.24) achieves a lower RMSE than the one trained with σ = 1.0, despite offering a stronger privacy guarantee. This non-monotonic trend suggests that moderate noise injection may improve generalization, consistent with prior work showing that carefully tuned DP noise can act as an implicit regularizer in deep learning models (Papernot et al., 2025; Bu et al., 2020). Similar effects have been reported in DP literature, where models trained with intermediate noise levels have outperformed low-noise counterparts in accuracy under certain optimization regimes (Abadi et al., 2016). Beyond σ = 3.0, performance appears to plateau, indicating robustness of the architecture to higher noise levels.

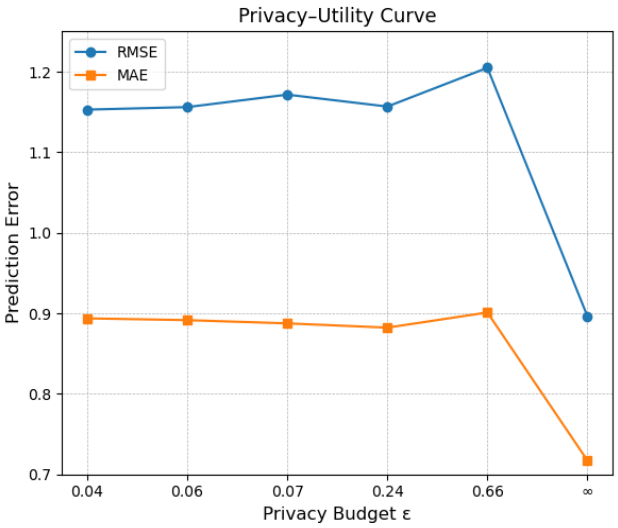

Figure 4. Prediction error results for differentially private Wide & Deep models RMSE and MAE versus privacy budget ε.

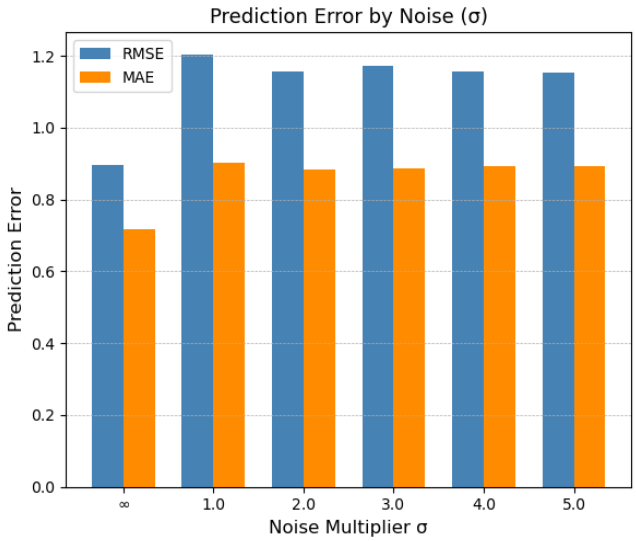

Figure 5. Prediction error results for differentially private Wide & Deep models RMSE and MAE versus noise multiplier σ, illustrating the privacy-utility relationship.

4.4. Discussion

These results confirm that integrating DP-SGD into a Wide & Deep recommender system is both feasible and effective for enforcing formal privacy guarantees while supporting the deployment of privacy-aware AI systems that handle sensitive user data. While increased noise generally leads to reduced accuracy, the degradation remains modest and predictable, particularly when $\varepsilon \leq 1.0$.

The non-monotonic pattern observed between $\sigma = 1.0$ and $\sigma = 2.0$ supports the hypothesis that moderate noise levels may enhance generalization. As noise increases further, model performance stabilizes rather than collapsing, indicating robustness to strong privacy constraints. This aligns with our expectation that the architecture's capacity for abstraction via dense embeddings mitigates the learning disruption introduced by DP mechanisms.

From a deployment perspective, these findings are significant. They demonstrate that strong user-level privacy, within practical privacy budgets such as $\varepsilon \leq 1.0$, can be achieved without compromising core recommendation performance. These results provide actionable guidance for designers of large-scale recommender pipelines: moderate DP noise levels preserve personalization accuracy while enforcing user-level privacy guarantees. This behavior illustrates that privacy constraints can be incorporated as part of algorithmic design rather than as an afterthought. In practice, this supports the adoption of DP-SGD in privacy-aware recommender systems deployed in practical settings. This is particularly relevant in compliance-driven environments governed by data protection regulations. Collectively, these insights contribute to the broader goal of building privacy-aware recommendation pipelines.

This suggests that, beyond a certain threshold, increasing noise has diminishing impact on utility. This pattern is consistent with prior observations in DP-SGD literature (Yu et al., 2019; Balle & Wang, 2018). This aligns with earlier findings showing reduced sensitivity to noise after convergence (Gopi et al., 2025).

In addition to overall accuracy, we also examined the distribution of RMSE values across different noise multipliers. As shown in Figure 6, the non-private model exhibits relatively low variance, but interestingly, the differentially

private models, especially those with σ = 2.0 and σ = 5.0, maintain tight and stable error distributions. This suggests that, despite the presence of noise, the Wide & Deep architecture exhibits robust generalization, with no major instability in prediction outcomes. The model appears capable of absorbing high noise levels without a corresponding increase in error variability, reinforcing its suitability for deployment in privacy-constrained environments.

While this study provides empirical insights into privacy-aware training for hybrid recommender models, it is limited to a single benchmark dataset and fixed architectural hyperparameters. In particular, reliance on a single dataset may limit the generalizability of the findings across different domains and data distributions.

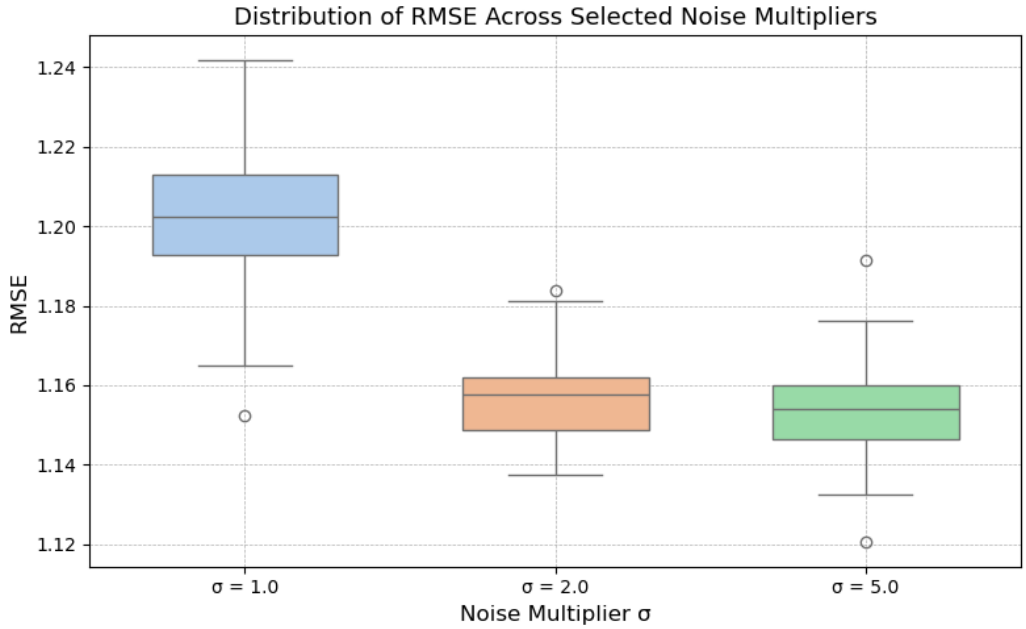

Figure 6. Boxplot showing the distribution of RMSE values across selected noise multipliers (σ = ∞, 1.0, 2.0, and 5.0). The low variance observed at higher noise levels indicates the model's robustness under strong privacy constraints

5. Conclusion and Future Work

This paper examined the privacy-utility trade-off in personalized recommendation by applying differential privacy to a Wide & Deep neural architecture. We implemented a differentially private training procedure using DP-SGD and evaluated model performance across varying noise levels on the MovieLens 1M dataset (Harper & Konstan, 2015). Both private and non-private models were compared using standard metrics (RMSE and MAE), and privacy guarantees were quantified using the Rényi differential privacy accountant.

The results show that differential privacy can be effectively integrated into deep hybrid recommenders while maintaining competitive accuracy. Notably, models trained with moderate noise (e.g., σ = 2.0, ε = 0.24) achieved better accuracy than those trained with lower noise, suggesting that noise injection may act as a form of regularization under certain conditions. Performance remained stable even at higher noise levels, indicating that the Wide & Deep architecture is not only privacy-compatible but also capable of maintaining trustworthy performance under stringent noise conditions.

This study provides an empirical foundation for the use of DP-SGD in structured recommender

settings. Future work may explore extensions such as adaptive noise schedules, personalization-aware privacy budgets, or integration with federated learning frameworks. Additional experiments on larger or domain-specific datasets may further validate the generalizability of the approach.

Overall, the findings support the feasibility of applying formal privacy guarantees in deep recommendation systems without substantially compromising personalization quality. By evaluating DP-SGD within a hybrid neural architecture, this study contributes to the development of practical privacy-preserving recommendation models.

Funding: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Project No. KFUXXXXXX).

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (pp. 308-318).

Balle and Y. X. Wang, “Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising,” 35th International Conference on Machine Learning, ICML 2018, vol. 1, pp. 678–692, May 2018, Accessed: Jul. 03, 2025. [Online]. Available: https://arxiv.org/abs/1805.06530v2

Bu, Z., Dong, J., Long, Q., & Su, W. J. (2020). Deep learning with gaussian differential privacy. Harvard data science review, 2020(23), 10-1162.

Calandrino, J. A., Kilzer, A., Narayanan, A., Felten, E. W., & Shmatikov, V. (2011, May). " You might also like:" Privacy risks of collaborative filtering. In 2011 IEEE symposium on security and privacy (pp. 231-246). IEEE.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2021). Extracting training data from large language models. In 30th USENIX security symposium (USENIX Security 21) (pp. 2633-2650).

Chaudhuri, K., Monteleoni, C., & Sarwate, A. D. (2011). Differentially private empirical risk minimization. Journal of Machine Learning Research, 12(3).

Chen, J., Zhang, H., He, X., Nie, L., Liu, W., & Chua, T. S. (2017, August). Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 335-344).

Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Shah, H. (2016, September). Wide & deep learning for recommender systems. In Proceedings of the 1st workshop on deep learning for recommender systems (pp. 7-10).

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006, March). Calibrating noise to sensitivity in private data analysis. In Theory of cryptography conference (pp. 265-284). Berlin, Heidelberg: Springer Berlin Heidelberg.

Fredrikson, M., Jha, S., & Ristenpart, T. (2015, October). Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security (pp. 1322-1333).

Gopi, S., Lee, Y. T., & Wutschitz, L. (2021). Numerical composition of differential

privacy. Advances in Neural Information Processing Systems, 34, 11631-11642.

Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. arXiv preprint arXiv:1703.04247.

Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis), 5(4), 1-19.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In Proceedings of the 26th international conference on world wide web (pp. 173-182).

Li, T., Song, L., & Fragouli, C. (2020, June). Federated recommendation system via differential privacy. In 2020 IEEE international symposium on information theory (ISIT) (pp. 2592-2597). IEEE.

McSherry, F., & Mironov, I. (2009, June). Differentially private recommender systems: Building privacy into the netflix prize contenders. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 627-636).

Mironov, I. (2017, August). Rényi differential privacy. In 2017 IEEE 30th computer security foundations symposium (CSF) (pp. 263-275). IEEE.

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In 2008 IEEE Symposium on Security and Privacy (sp 2008) (pp. 111-125). IEEE.

Papernot, N., Song, S., Mironov, I., Raghunathan, A., Talwar, K., & Erlingsson, Ú. (2018). Scalable private learning with pate. arXiv preprint arXiv:1802.08908.

Ran, X., Wang, Y., Zhang, L. Y., & Ma, J. (2022). A differentially private matrix factorization based on vector perturbation for recommender system. Neurocomputing, 483, 32-41.

Ran, X., Ye, Q., Hu, H., Huang, X., Xu, J., & Fu, J. (2024, May). Differentially private graph neural networks for link prediction. In 2024 IEEE 40th International Conference on Data Engineering (ICDE) (pp. 1632-1644). IEEE.

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks against machine learning models. In 2017 IEEE symposium on security and privacy (SP) (pp. 3-18). IEEE.

Tensorflow community, "tensorflow/privacy: Library for training machine learning models with privacy for training data," GitHub. Accessed: Jun. 08, 2025. [Online]. Available: https://github.com/tensorflow/privacy

Wang, Y. X., Balle, B., & Kasiviswanathan, S. P. (2019, April). Subsampled rényi differential privacy and analytical moments accountant. In The 22nd international conference on artificial intelligence and statistics (pp. 1226-1235). PMLR.

Wu, C., Wu, F., Lyu, L., Qi, T., Huang, Y., & Xie, X. (2022). A federated graph neural network framework for privacy-preserving personalization. Nature Communications, 13(1), 3091.

Ye, Q., Hu, H., Meng, X., & Zheng, H. (2019, May). PrivKV: Key-value data collection with local differential privacy. In 2019 IEEE Symposium on Security and Privacy (SP) (pp. 317-331). IEEE.

Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. Advances in neural information processing systems, 32.