Performance Prediction for Learning to Rank Systems

Aaron Goh
University of Utah

UUCS-21-007

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

22 April 2021

Abstract

Learning to Rank has been an intensively studied topic within the Information Retrieval community. Among its top applications is with enabling Search Engines to identify relevant documents from a query input. With Search Engines being the primary engine behind on-demand access to information, it is important for Learning to Rank algorithms to return trustworthy documents to users as poor results have the potential to greatly misinform individuals. In this thesis, we propose utilizing machine learning models to predict the performance of popular existing learning to rank algorithms Ranking SVM, LambdaMart, and RankNet. This model helps existing web search engines augment successful evaluation of rank lists. Expressing the viability of this solution, empirical results for Feedforward, LSTM, GRU, Transformer, and other model variants have been collected and evaluated based on the C14-Yahoo! Learning to Rank Challenge Data Set.

Performance Prediction for Learning to Rank Systems

By

Aaron Goh

A Senior Thesis

Submitted to the Faculty of The University of Utah in

Fulfillment of the Requirements for the

Degree of Bachelor of Science

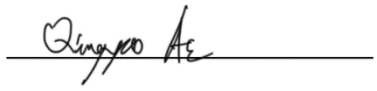
in Computer Science

School of Computing

University of Utah

April 21, 2021

Approved:



Qingyao Ai, PhD

Thesis Faculty Supervisor

H. James de St. Germain, PhD

Director of Undergraduate Studies

School of Computing

Mary Hall

Director of

School of Computing

Abstract

Learning to Rank has been an intensively studied topic within the Information Retrieval community. Among its top applications is with enabling Search Engines to identify relevant documents from a query input. With Search Engines being the primary engine behind on-demand access to information, it is important for Learning to Rank algorithms to return trustworthy documents to users as poor results have the potential to greatly misinform individuals. In this thesis, we propose utilizing machine learning models to predict the performance of popular existing learning to rank algorithms Ranking SVM, LambdaMart, and RankNet. This model helps existing web search engines augment successful evaluation of rank lists. Expressing the viability of this solution, empirical results for Feedforward, LSTM, GRU, Transformer, and other model variants have been collected and evaluated based on the *C14 – Yahoo! Learning to Rank Challenge Dataset*.

1 Introduction

Search engines, backed by Learning to Rank (LTR) models, have yielded high value to the general population with on-demand information. With these search engines, a query can be input from a user where a list of relevant documents will be identified from a Learning to Rank model. However, in recent times, ambiguous queries looking for information in less explored domains have returned untrustworthy results. In one specific case, the student Wei Zexi received a rare form of cancer and queried Baidu to find a cure. However, less proven methods rose to the top of the documents which led to Wei Zexi taking poor medical advice - eventually leading to his death [3].

This has been a well-known problem in IR introducing the concept of Robustness which reflects how well an IR system performs for poorly performing queries. In extension to this problem, we propose utilizing the tree-based models and artificial neural network (ANN) models to predict the performance of existing LTR model performances which may better inform the credibility of the IR system. In specific, we consider utilizing the Light Gradient Boosting Machine (LightGBM [46]) and such as Multilayer Perceptron, LSTM, GRU, and Transformer networks. We also consider utilizing self-attention on top of existing models such as LSTM and GRU which has empirically been shown to improve performance in semantic modeling in comparison to without self-attention.

As a final evaluation for model performance based on R-Score, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) between predicted query performance and actual performance, this thesis evaluates resultant QPP scores by clusters generated from k-means determine if the QPP model struggles or excels in predicting poor or strong queries.

2 Background and Related Work

Learning to Rank represents techniques that attempt to solve ranking problems using Machine Learning Models with features represented in document-query pairs. In existing implementations, Learning to Rank models have been successfully applied to fields such as recommendation in context of social media engagement [4] and document retrieval [5].

Previous focus of IR research in the 1980s / 1990s for ranking problems in IR is the traditional non-learning approach such as BM25 [18] and language models [19] which effectively try to deduce the relevance as probabilistic ranking based on frequency of document terms matching with query terms where parameters for the models are learned heuristically. However, this is applied to ad-hoc retrieval where the documents and queries are both textual.

2.1 Learning to Rank Approaches

Learning to Rank models return a score which indicates the relevance for documents in a query. In comparison to traditional ML problems, LTR's primary objective is to create a rank list from the list of documents rather than a singular output representing a class or numeric representation.

There are three core methods to implementing LTR models - pointwise, pairwise, and listwise approaches. The approaches define different input and output spaces, use different hypotheses, utilize different loss functions between various models.

2.1.1 Pointwise Method

In pointwise, the ranking problem is similar to most traditional machine learning models closer to a classification or regression problem by viewing each input document as a learning instance and directly predicting a relevance score seen in methods such as McRank [8]. Effectively, each score for each document is independent of each other and has been criticized for not really maximizing a ranking objective [22].

2.1.2 Pairwise Method

For the pairwise method, the algorithm takes two documents at a time and attempts to rank them relative to each other to eventually generate the final list seen in methods such as Ranking SVM, LambdaMart, and RankNet [7]. In general, for pairwise modeling, the learning instances are document pairs $f(x_i) - f(x_j)\ \forall x_i, x_j \in X$. In the case of Ranking SVM, these instances are then used to solve classification with support vector machines following the classical quadratic optimization problem $\min_{\vec{w}} M(\vec{W}) = \frac{1}{2}||\vec{w}||^2 + C\sum_{i=1}^{t}\xi_i$.

LambaMart is also a pairwise approach which combines LambdaRank and MART (Multiple Additive Regression Trees) which considers using gradient boosted decision trees from a cost function derived from LambdaRank which effectively does an empirical optimization of an evaluation metric such as NDCG. RankNet, similar to LambdaMart, also considers gradient boosted decision trees but utilizes Cross Entropy as a loss function and Gradient Descent as the algorithm to train a Neural network [21].

This approach's advantage is that classification can be directly applied and thus handled as a traditional machine learning model. However, this means that the loss function is

minimizing the relevance between document pairs rather than in consideration of the complete rank list.

2.1.3 Listwise Method

In the listwise approach, the learning instance takes the entire list of documents associated to a query as a singular input and maximizing the ranking metrics for each feature vector seen in methods such as ListNet [9] and ListMLE [10]. Listwise approaches take the output of some ranking function per document utilized as a feature vector then minimized or maximized from a listwise loss function. Afterwards, the scores are sorted by descending document which is now viewed as Learning to Rank. Listwise methods also often utilize probabilistic views of the scores for varying permutations [9].

2.2 Neural-based & Tree-based Models

Current applications of neural-networks so far have been focused on handling more complex ad-hoc tasks and improving on NLP-based approaches in IR and image searching. CNNs have been used to map sentence models semantic matching between query-document pairs to determine ranking [24].

2.2.1 Feedforward

Feedforward networks also known as Multi-layer Perceptron (MLP) are one of the most popular and widely used neural networks. In LTR, this has been used as an approximation model for LambdaMart and other gradient tree-based model with the benefit of performance given that the tree's evaluation cost is $O(Td)$ for a forest size T and depth d [28].

2.2.2 Recurrent Neural Networks

Long Short-Term Memory (LSTM) is a model improving on vanilla RNNs by addressing the vanishing gradient issue utilizing "gates." Due to the presence of the forget gate combined with the additive property of cell state gradients, the network is able to update parameters and prevent the error gradients from vanishing [31].

The Gated Recurrent Unit (GRU) is an increasingly popular variant of LSTM. In comparison to LSTM, it does not utilize cell states and instead uses hidden states to transfer information [34] between units. Generally, GRU and LSTM solve the same tasks however, GRU has been known to be a bit more computationally efficient [34]. Among the applications of LSTM and GRU within LTR is question-answer topic cluster ranking [44] and in designing LTR products based on product reviews in hierarchal networks [45].

2.2.3 Attention Mechanism

Prior to the attention-mechanism, sequence-to-sequence models have been primarily utilized with the issue of maintaining longer term sequence information. While, Attention so far has been applied primarily to NLP tasks and in learning to rank, typically centered around ad-hoc retrieval, other sequential modeling such as stock price predictions have also utilized attention models to enhance performance [36].

Self-attention has also been applied in conjunction with LSTM and GRU models interacting with either the raw input before passing the attention vector into the network or utilizing the hidden state of LSTM and GRU model as inputs to the attention layer to improve sentiment analysis in long sentences [35] effectively focusing on different parts of the sequence. The attention layer produces the attentional hidden state as follows: $\widetilde{h_t} = \tanh\left(W_c[c_t; h_t]\right)$ and

is fed through a softmax layer representing the predictive distribution $p(y_t|y_{<t}, x) = softmax(W_s \tilde{h}_t)$.

One of the schemes known as the global attention mechanism [37] considers all hidden states of the encoder when deriving the context vector c_t which varies from other schemes such as the local attention that focus on subsets of the hidden states. However, generally, it is set to be $c_t = \sum_{j \in s'} \alpha_{tj} h_j$. Reviewing the global attention scheme briefly, we create an alignment vector α_t whose size is equal to the number of time steps in the source side, which is derived by comparing the current target hidden state h_t with each source hidden state $\bar{h}_s$

$$\alpha_t = align(h_t, \bar{h}_s) = \frac{\exp\left(score(h_t, \bar{h}_s)\right)}{\sum_{s'} \exp\left(score(h_t, \bar{h}_{s'})\right)}$$

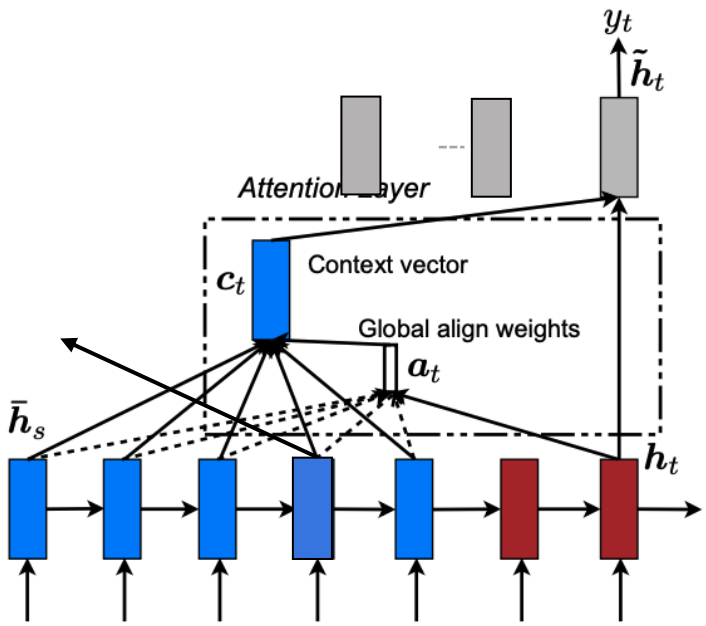


Figure 1. Global Attention Layer adapted from [37]

2.2.4 Transformer

The Transformer model has recently been developed on top of the attention mechanism which introduces an encoder-decoder model which has primarily been applied to language translation [38].

2.2.5 Activation Functions

At the end of each neural network, an activation function is present to determine the output of each neuron. In development of our deep learning models, we consider the following activation functions: tanh, sigmoid, softmax, and relu.

As a brief recap, $\tanh(x)$ is defined as $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ where x is an element within the vector represents the state of a neuron. $\tanh(x)$ is also zero-centered between -1 and 1 but suffers with the vanishing gradient problem and is computationally expensive.

$sigmoid(x)$ is defined as $\sigma(x) = \frac{1}{1+e^{-x}}$ where x is an element within the vector representing the state of a neuron. Similar to $\tanh(x)$, this has a vanishing gradient problem and is computationally expensive however, the outputs are not zero-centered but are bounded between 0 and 1.

$softmax(x)$ is defined as $\sigma(x) = \frac{e^x}{\sum_{j=1}^{K} e^x}$ where K is the number of outputs for the layer. In effect, this gives the probability distribution of different output candidates given that we guarantee the outputs sum of this layer sum up to 1.

$relu(x)$ is defined as $f(x) = \max(0, x)$ which is differentiable at all points but 0. This is the most computationally simple with fast convergence [27]. It has also become one the most popular activation function activation functions as its behavior is also close to linear while being

non-linear [26]. However, when the majority of inputs are ≤ 0, the gradient effectively vanishes which prevents effective learning [25, 26].

2.3 Evaluating IR Model Performance

For LTR systems, among the common metrics for measuring the performance of a query-document list are the Normalized Discount Cumulative Gain (NDCG), Precision, Reciprocal Rank (RR), and Expected Reciprocal Rank (ERR).

2.3.1 Normalized Distributed Cumulative Gain (NDCG)

$NDCG@k$ is a measure evaluating the top k documents of a rank list that considers multi-leveled relevance labels. Specifically, it is defined as

$$NDCG_k = \frac{DCG_k}{IDCG_k}$$

where $DCG_k = rel_1 + \frac{\sum_{i=2}^{k} rel_i}{\log_2(i+1)}$ with rel_i being the relevance score at position i and $IDCG_k$ is the idealized version of DCG_k where each rel_i is sorted creating the best possible rank list [11].

Alternatively, we can define $DCG_k = \sum_{i=1}^{k} \frac{2^{rel_i}-1}{\log_2(i+1)}$. In this thesis, we utilize the first scheme instead of this one which puts more weight on relevant queries. Generally, this is thought of to mimic user behavior where we highly emphasize only the top few documents with diminishing value.

2.3.2 Precision

$P@k$ is a measure evaluating the top k documents of a rank list however, it only considers binarized relevance labels (i.e., relevant and non-relevant). It is defined as

$$P@k = \frac{1}{k}\sum_{i=1}^{k} r_i$$

where r_i is the binarized relevance score at position i.

2.3.3 Expected Reciprocal Rank (ERR)

ERR_k is a measure evaluating the top k documents of a rank list which can consider multi-level relevance scores. It is defined as

$$ERR_k = \sum_{i=1}^{k} \frac{1}{i} * \prod_{j=1}^{k-1}(1 - R_j) * (R_i)$$

where $R_i = \frac{2^{rel_i - 1}}{2^{\max(\text{rel})}}$ which represents the probability of relevance for the document at i. This is thought of as a cascading model which considers that users view documents from top to bottom and that there is an associated probably at each position i that a user will be satisfied [20].

2.3.4 Reciprocal Rank (RR)

RR_k is a measure evaluating the top k documents of a rank list which can consider multi-level relevance scores. It is defined as

$$RR_k = \sum_{i=1}^{k} \frac{1}{R_i}$$

where R_i is the rank of the highest relevant document within the query.

2.4 Query Performance Prediction (QPP)

One of the known problems within LTR is the difference in performance between poorly performing queries and strongly performing queries. Within ad-hoc retrieval tasks where the data is often viewed as an NLP task between documents to predict relevance, there have been previous developments for query performance prediction in pre-retrieval query performance prediction focusing on features generated from the query string to further filter performance of models [15].

In existing research to better improve QPP, there have been prediction models such as Weighted Information Gain (WIG) which is applied to content-based and Named-Page (NP) finding queries [6]. Clarity score has also been utilized as a measure to estimate the performance for a query language model which effectively suggests high coherence measured as the KL-divergence between the query and the document collection as a unigram language model [16].

In [12], there have been models utilizing statistical decision theory as post-retrieval predictors. In comparison, this thesis explores the problem using Learning to Rank models as the retrieval model rather than ad-hoc retrieval models which also enables general evaluation for feature vectors which is limited in content-based or Named-Page finding queries. However, predictions of pre-retrieval query performance prediction utilizing machine learning on traditional web search features has not been explored yet to the author's knowledge.

3 Methods

In the methods section, we first describe the pre-processing of the data necessary to formulate the problem and then describe the models used to predict the QPP score. Lastly, we consider robustness and describe the methodology for determining model performance.

3.1 Data Information

The data used is the *C14 – Yahoo! Learning to Rank Challenge Dataset* [6]. The features used are not explicitly labeled however are generically divided into the categories: Web Graph, Document Statistics, Document classifier, Query, Text match, Topical Matching, Click, External references, and Time. The dataset contains different feature types including binary, count, and continuous considerations.

The relevance labels per each document are labeled from 0 (for bad) and 4 (for perfect).

Table 1. Yahoo LTRC Set 1 Information

Set 1	Train	Val	Test
Queries	19,944	2,994	6,983
# Documents	473,134	71,083	165,660
# Features	700	700	700

In specific, the dataset follows the traditional LTR problem formulation as below:

$$Q = \{q^1, q^2, \dots, q^m\}$$

where Q is the set of queries and m is the number of queries.

Each query q_i has a list of documents

$$d^i = \{d_1^i, d_2^i, \dots, d_{n(i)}^i\}$$

where $n(i)$ is the number of documents for the respective query.

We first collect rank lists from various algorithms – Ranking SVM from SVM Joachim's SVM light algorithm [39] and LambdaMart / RankNet algorithms from the Lemur Project's RankLib [40]. We calculate the NDCG@5, NDCG@10, RR@5, RR@10, ERR@5, ERR@10, Precision@5, and Precision@10 from each rank list created from each algorithm. Noticeably, for Precision, we binarize the relevance labels considering values {3, 4} as 1 (for good) and {1, 2, 3} as 0 (for bad).

Prior to training, we ensure that the dataset is converted to a dense version where all missing features in each document are set to 0. To simplify dimensionality constraints, for each query-document set, we only consider the top thirty documents output from the rank list which is the average number of documents per query. We implement the various deep neural networks (DNNs) and consider LightGBM [46] as the baseline decision tree to compare results.

3.2 Models

For each ANN model, we utilize the hyperband [42] hyperparameter tuning model and in this thesis, we only present the best model results. Generally, for each model, we utilize the standard stochastic gradient descent optimization considering learning rates $[1e-1, 1e-2, 1e-3, 1e-4, 1e-5]$ and utilize the mean squared error loss function.

3.2.1 Feedforward

The feedforward network utilized follows the standard formulation where the value at each neuron is denoted as $a^{(l)} = \sigma(Wa^{l-1} + b)$ where W denotes a matrix of weights, $a^{(l)}$ represents some activation at layer (l), b is some bias, and $\sigma(x)$ is some activation function. During optimization, we consider some cost function C and update the gradient that calculates the weight at each layer as $w^{(l)} = w^{(l)} - \eta * \frac{\partial C}{\partial w^{(l)}}$ where η is the learning rate and $\frac{\partial C}{\partial w^{(l)}}$ is the cost gradient with respect to the weight. Backpropagation occurs at the end where we go backwards from the output layer towards the input layer and update the gradients for the weights and biases [27]. In our case, we also utilize mini batching with batch size 10. Backpropagation occurs only after seeing m training instances [30].

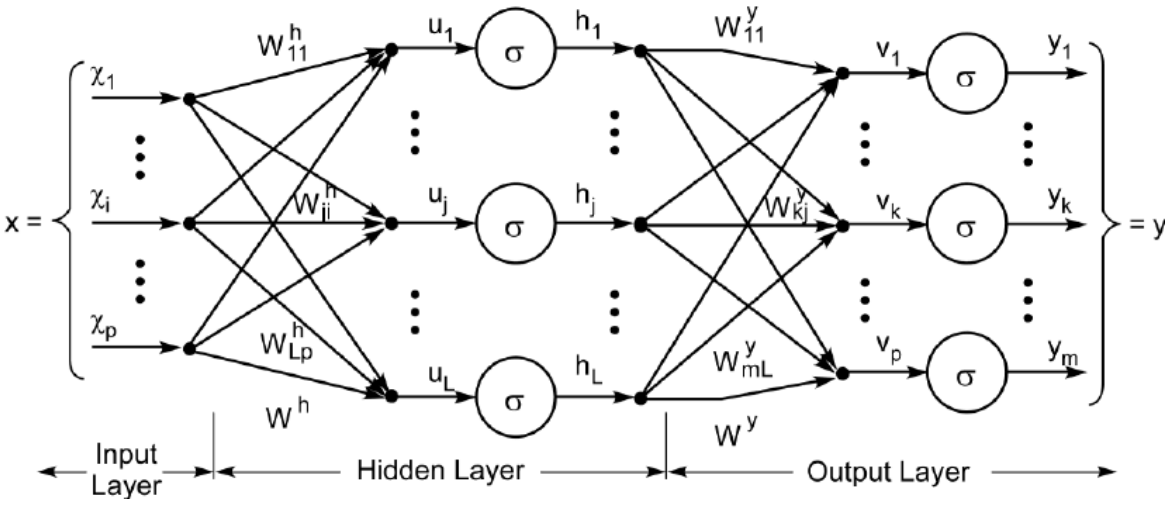


Figure 2. Feedforward Diagram from [33]

Considering the Feedforward network implementation, we implement a naive implementation where the query-document lists are flattened as a singular input and missing documents are padded with 0s for the feature. For hyperparameter training, we consider training

on three dense layers evaluated between dimensions $[32 - 768]$ with step size 64 before passing through a final output neuron. For each neuron, we consider activation functions $softmax$, $relu$, and $sigmoid$.

3.2.2 Attention

We utilize the attention mechanism (Refer to Figure 1.) as a singular layer where the inputs consider the initial dataset directly to predict the query performance. As a scoring mechanism, we use Luong's multiplicative style defined as $score(h_t, \bar{h}) = h_t^T W_\alpha \overline{h_s}$. In preparing the data for Attention, we utilize one-hot encoding to include positional embedding for the rank of each document input per query.

3.2.3 Long Short-Term Memory (LSTM)

For the LSTM network, we consider the top ten documents from the rank list and set the batch size to ten setting each document-query list as a single batch. The model is implemented without variation on the standard LSTM cell which is defined as follows:

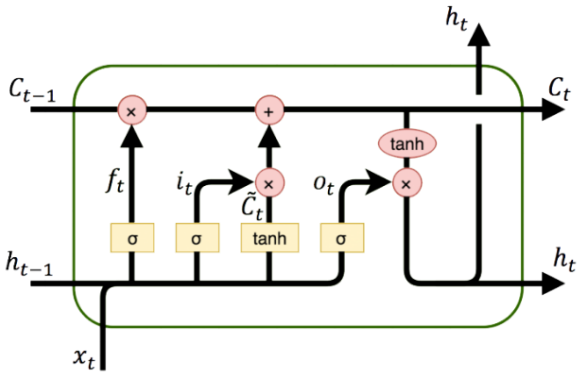


Figure 3. LSTM Model Diagram taken from [32]

$$f_t = \sigma(W_{fh}\, h_{t-1} + W_{fx} x_t + b_f) \quad (f_t\text{: forget gate}) \quad (1)$$

$$i_t = \sigma(W_{ih} h_{t-1} + W_{ix} x_t + b_i) \quad (i_t\text{: input gate}) \quad (2)$$

$$\tilde{C_t} = \tanh\left(W_{\tilde{C}h} h_{t-1} + W_{\tilde{C}x} x_t + b_{\tilde{C}}\right) \quad (\tilde{c_t}\text{: cell input}) \quad (3)$$

$$C_t = C_{t-1} + i_t \odot \tilde{C_t} \quad (c_t\text{: cell state}) \quad (4)$$

$$o_t = \sigma(U_{oh} h_{t-1} + W_{ox} x_t + b_o) \quad (o_t\text{: output gate}) \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (h_t\text{: hidden state}) \quad (6)$$

We augment the LSTM network by inputting the hidden states at the last LSTM cell into a self-attention network utilizing Luong's multiplicative style defined as $score(h_t, \bar{h}) = h_t^T W_\alpha \overline{h_s}$ in a global attention scheme per time series. Specifically, we take the hidden cell states $(h_1, h_2, \dots, h_n) \in \mathbb{R}^{d_x}$ of the LTM units and pass each h_i into an attention layer. We consider h_t to be the final hidden state of the networks with h_s. In effect, this is an encoder-decoder model where the encoder is encapsulated through the hidden states of the LSTM or GRU networks and the decoder effectively utilizes the context vector and target hidden state to decode a target y_t.

For hyperparameter training, we consider two LSTM layers with dimensions between $[32 - 768]$ with step size 64 with a final dense layer considered with activation functions $softmax$, $relu$, and $sigmoid$. When considering the global attention mechanism (Refer to Figure 1.) as an additional layer past the LSTM cells, we consider the attention vector dimensions between $[32 - 768]$ with step size 64 before passing through the output neuron where we consider the activation functions $softmax$, $relu$, and $sigmoid$.

3.2.4 Gated Recurrent Unit (GRU)

Similar to the LSTM network, we consider the top ten documents from the rank list and set the batch size to ten setting each document-query list as a single batch. The model is implemented without variation of the traditional GRU cell defined as follows:

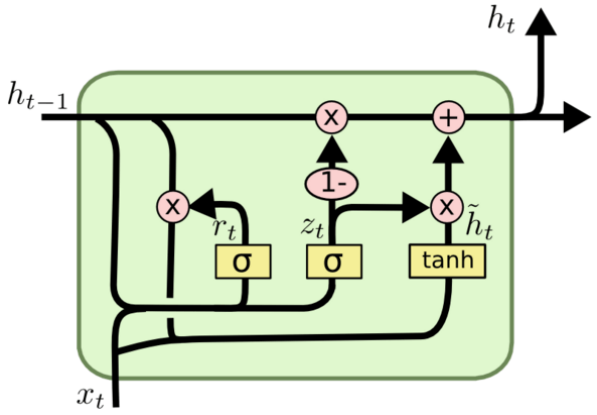


Figure 4. GRU Model taken from [43]

$$z_t = \sigma(W_{zh} \cdot h_{t-1} + W_{zx} \cdot x_t + b) \quad (7)$$

$$r_t = \sigma(W_{rh} \cdot h_{t-1} + W_{rx} \cdot x_t + b) \quad (8)$$

$$\tilde{h_t} = \tanh\,(W_{rh} \cdot r_t h_{t-1} + W_{rx} \cdot x_t + b) \quad (9)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h_t} \quad (10)$$

We augment the GRU network by inputting the hidden states at the last GRU cell into a self-attention network utilizing Luong's multiplicative style defined as $score(h_t, \bar{h}) = h_t^T W_\alpha \overline{h_s}$ in a global attention scheme per time series. Specifically, we take the hidden cell states $(h_1, h_2, \dots, h_n) \in \mathbb{R}^{d_x}$ of the GRU units and pass each h_i into an attention layer. We consider h_t to be the final hidden state of the networks with h_s. In effect, this is an encoder-decoder model

where the encoder is encapsulated through the hidden states of the LSTM or GRU networks and the decoder effectively utilizes the context vector and target hidden state to decode a target y_t.

3.2.5 Transformer

Within the transformer network, typically positional encoding occurs however, for the context of the dataset used, the Yahoo Dataset does not reveal explicit features to prevent reverse engineering and is a continuous feature set with no explicit embedding information. For simplicity, in this usage of the transformer, we do not consider the positional encoding step meaning we specifically only consider the initial encoder layer described in [38] without the embedding step provided our dataset contains continuous values.

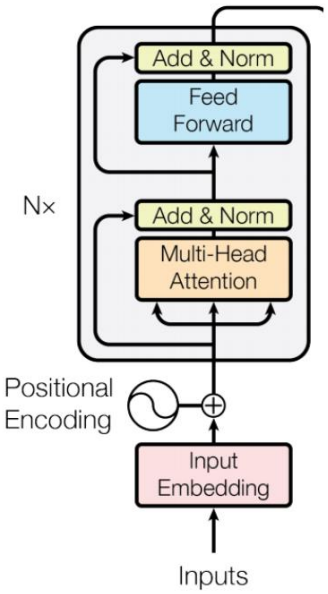


Figure 5. Transformer Model From [38]

In the transformer model, the multi-head attention is utilized to parallelize training of attentions applied within each embedded space. Specifically, it divides the space of features by

h, the number of heads, and effectively splits the attention to focus on different parts of the sequence.

The self-attention mechanism that occurs in the multi-head attention is formulated as

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

where d_k is the length of the K, V matrices where the Q, K, and V matrices are equivalent to the sequence input. However, given multiple heads, $d_k = d_v = \frac{d_{model}}{h}$.

With the Multi-Head attention, this ends up as

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O$$

with $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$

with parameter weight matrices $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$, and $W^O \in \mathbb{R}^{hd_v \times d_{model}}$ trained from typically single feedforward neuron.

In contrast to the traditional transformer model, we only consider the encoder side of the model excluding the decoder side of the model which effectively makes the model utilize the multi-head attention mechanism and a feedforward network with normalization between layers. Furthermore, similar to the attention layer, we utilize one-hot encoding to include positional embedding for the rank of each document input per query as part of the input rather than utilizing positional encoding given no properties available from embedding.

For the hyperparameter tuning, we consider utilizing either 8 or 16 heads and pad any remaining dimensions with $0s$ to ensure the feature set is split appropriately between the different heads. Provided the feedforward layers in between the self-attention layers and the output, we consider the activation functions $tanh$, $softmax$, $relu$, and $sigmoid$.

In the case without the feedforward network, this output is flattened and passed through a single output neuron considering the activation functions $softmax$, $relu$, and $sigmoid$. Otherwise, we pass the transformer output through an MLP with three layers trained with the same considerations as the feedforward model.

3.2.6 LightGBM

In the LightGBM model, we utilize the regressor version and flatten the query-document pair features prior to training as input to the model. For hyperparameter tuning, we consider number of leaves between $[6, 50]$ with step size 4, minimum child samples between $[100, 500]$ with step size 20, minimum child weights $[1e-5, 1e-3, 1e-2, 1e-1, 1]$, a maximum depth between $[10, 100]$ with step size 10, a regularization term α between $[1e-5, 1e-3, 1e-2, 1e-1, 0, 1]$, and a regularization term λ between $[1e-5, 1e-3, 1e-2, 1e-1, 0, 1]$.

3.3 Robustness Measurement Through Clustering

As part of a first step for robustness evaluation, we differentiate highly performing queries from poorly performing queries [16]. To do this, we utilize K-means clustering running for $n = 5$ to separate predictions output from the model. Utilizing different clusters, we consider whether the model performs better or worse towards strongly performing queries or poorly performing queries.

4 Results

4.1 Precision

4.1.2 Precision@5

Considering the results for Precision@5, the LightGBM model still outperforms the ANNs with the highest $R2$ score and lower $RMSE$ scores across each different underlying LTR algorithms. While overall predictability is scattered by an average deviation around .3 *[Table 1, Table 2, Table 3]*, this is still potentially useful as a rough augmentation onto existing document query features.

Table 1. RankingSVM - Precision@5

Precision@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1050	0.1090	0.1098	0.1048	0.0958	0.1049	0.1010	0.1088	0.1067
R2 Score:	0.3201	0.2941	0.2887	0.3213	0.3797	0.3206	0.3459	0.2958	0.3088
RMSE:	0.3240	0.3302	0.3314	0.3237	0.3095	0.3239	0.3178	0.3298	0.3267
MAE:	0.4976	0.5200	0.5064	0.4919	0.4884	0.5077	0.4938	0.5041	0.5172

Table 2. LambdaMart - Precision@5

Precision@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1115	0.1081	0.1124	0.1078	0.1022	0.1147	0.1087	0.1128	0.1152
R2 Score:	0.3415	0.3617	0.3363	0.3636	0.3965	0.3230	0.3580	0.3341	0.3196
RMSE:	0.3340	0.3288	0.3353	0.3283	0.3197	0.3386	0.3298	0.3358	0.3395
MAE:	0.5108	0.5054	0.5196	0.5118	0.5048	0.5243	0.5139	0.5144	0.5235

Table 3. RankNet - Precision@5

Precision@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1001	0.1018	0.1016	0.0983	0.0928	0.1019	0.1005	0.1035	0.1028
R2 Score:	0.2803	0.2680	0.2694	0.2930	0.3323	0.2674	0.2774	0.2561	0.2606
RMSE:	0.3164	0.3191	0.3187	0.3136	0.3047	0.3192	0.3170	0.3216	0.3207
MAE:	0.4958	0.4906	0.4912	0.4866	0.4809	0.4990	0.4913	0.4884	0.5042

Furthermore, when taking consideration of clusters *[Appendix 6.1.1, Figure 6, Figure 7, Figure 8, Figure 9, Figure 10]*, we notice that LightGBM outperforms all other metrics in terms of ability to predict considering the $R2$ score and that in general, the models typically perform better for higher predicted scores rather than poorly performing queries.

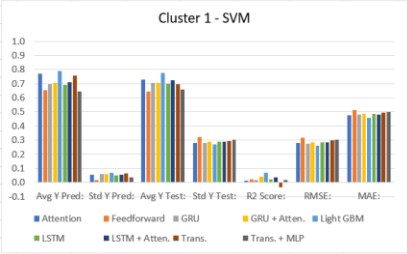


Figure 6a. Precision@5 – SVM – Cluster 1

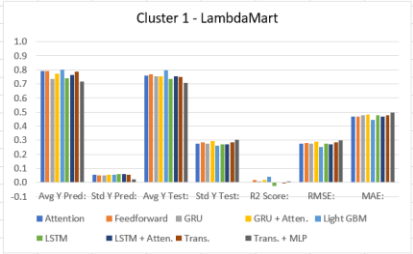


Figure 6b. Precision@5 – LambdaMart – Cluster 1

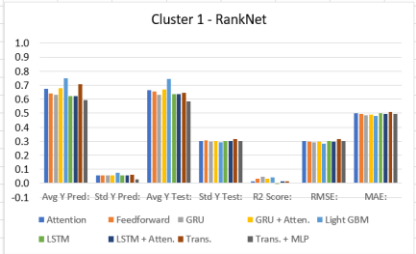


Figure 6c. Precision@5 – RankNet – Cluster 1

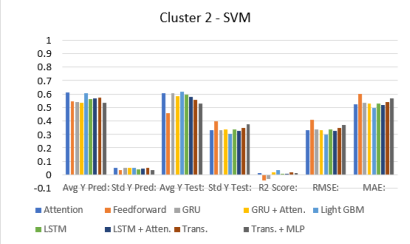


Figure 7a. Precision@5 – SVM – Cluster 2

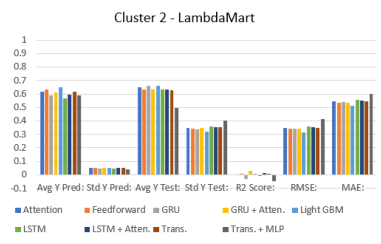


Figure 7b. Precision@5 – LambdaMart – Cluster 2

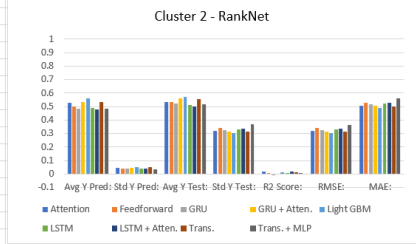


Figure 7c. Precision@5 – RankNet – Cluster 2

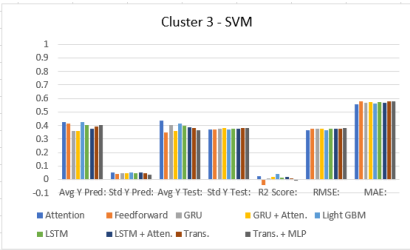


Figure 8a. Precision@5 – SVM – Cluster 3

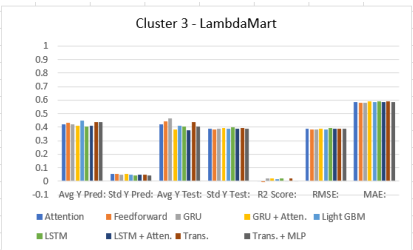


Figure 8c. Precision@5 – RankNet – Cluster 3

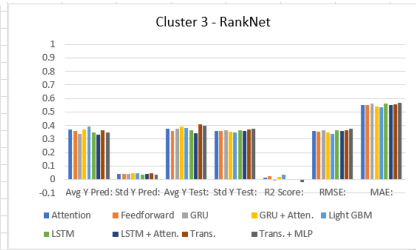


Figure 8b. Precision@5 – LambdaMart – Cluster 3

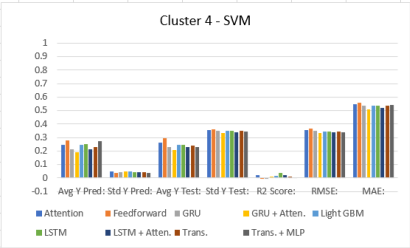


Figure 9a. Precision@5 – SVM – Cluster 4

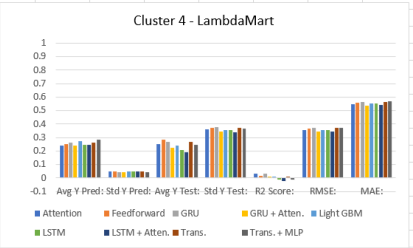


Figure 9b. Precision@5 – LambdaMart – Cluster 4

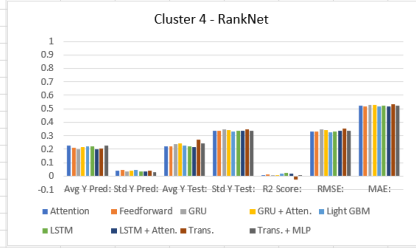


Figure 9c. Precision@5 – RankNet – Cluster 4

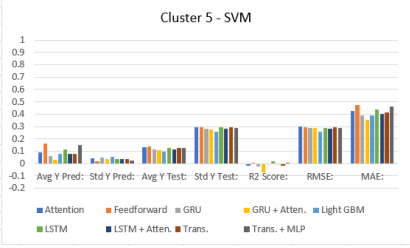


Figure 10a. Precision@5 – SVM – Cluster 5

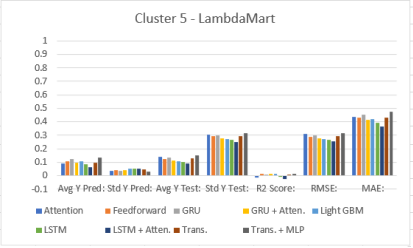


Figure 10c. Precision@5 – RankNet – Cluster 5

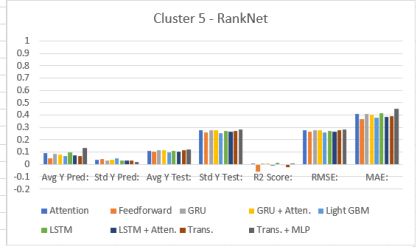


Figure 10b. Precision@5 – LambdaMart – Cluster 5

4.1.2 Precision@10

In the results for Precision@10, results are similar where the LightGBM model still outperforms the ANNs with the highest $R2$ score and lower $RMSE$ scores across each different underlying LTR algorithms *[Table 4, Table 5, Table 6]*. The overall average error is slightly higher within these models which may be a factor of the number of documents considered *[Table 1, Table 2, Table 3, Table 4, Table 5, Table 6]*.

Table 4 - RankingSVM - Precision@10

Precision@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1288	0.1259	0.1321	0.1255	0.1193	0.1277	0.1268	0.1327	0.1808
R2 Score:	0.2911	0.3073	0.2729	0.3095	0.3437	0.2972	0.3021	0.2698	0.0051
RMSE:	0.3589	0.3548	0.3635	0.3542	0.3454	0.3574	0.3561	0.3643	0.4252
MAE:	0.5381	0.5232	0.5491	0.5354	0.5216	0.5412	0.5421	0.5407	0.6244

Table 5 - LambdaMart - Precision@10

Precision@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1324	0.1356	0.1356	0.1301	0.1224	0.1326	0.1298	0.1476	0.1913
R2 Score:	0.3080	0.2913	0.2913	0.3198	0.3601	0.3068	0.3214	0.2287	-0.0001
RMSE:	0.3639	0.3682	0.3682	0.3607	0.3499	0.3642	0.3603	0.3841	0.4374
MAE:	0.5470	0.5596	0.5596	0.5399	0.5277	0.5493	0.5391	0.5365	0.6380

Table 6 - RankNet - Precision@10

Precision@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1265	0.1239	0.1275	0.1252	0.1180	0.1264	0.1284	0.1327	0.1705
R2 Score:	0.2580	0.2731	0.2519	0.2654	0.3077	0.2587	0.2469	0.2219	-0.0002
RMSE:	0.3557	0.3520	0.3571	0.3539	0.3435	0.3555	0.3583	0.3642	0.4129
MAE:	0.5413	0.5319	0.5446	0.5384	0.5203	0.5318	0.5493	0.5419	0.6118

Based on the clustering results *[Appendix 6.1.2, Figure 11, Figure 12, Figure 13, Figure 14, Figure 15]*, similar to Precision@5 *[Appendix 6.1.1, Figure 6, Figure 7, Figure 8, Figure 9, Figure 10]*, the performance of predictions decrease as predicted Precision@10 goes lower. In consideration of model performance in each cluster, LightGBM outperforms the ANN in each cluster.

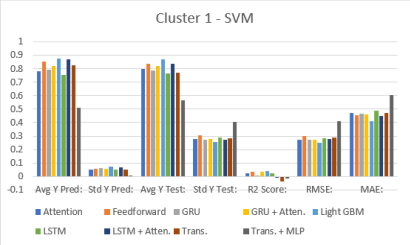


Figure 11a. Precision@10 – SVM – Cluster 1

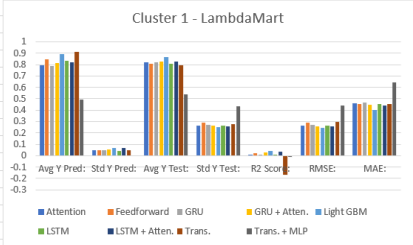


Figure 11b. Precision@10 – LambdaMart – Cluster 1

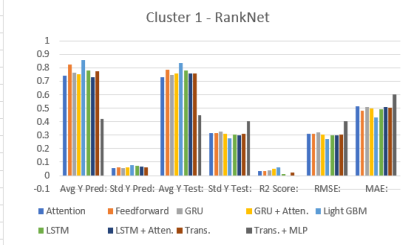


Figure 11c. Precision@10 – RankNet – Cluster 1

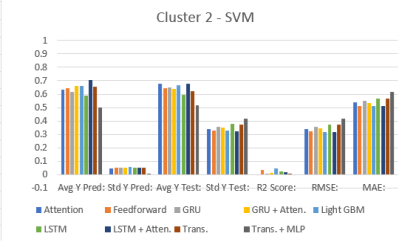


Figure 12a. Precision@10 – SVM – Cluster 2

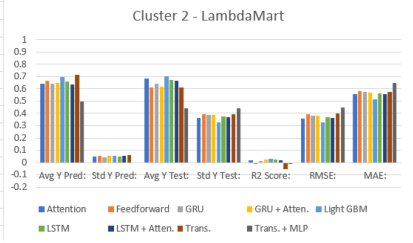


Figure 12b. Precision@10 – LambdaMart – Cluster 2

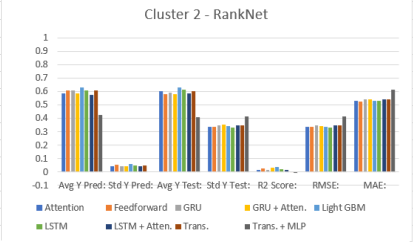


Figure 12c. Precision@10 – RankNet – Cluster 2

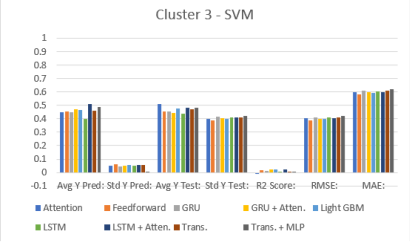


Figure 13a. Precision@10 – SVM – Cluster 3

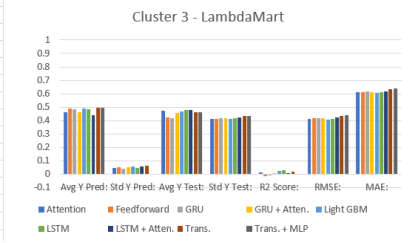


Figure 13b. Precision@10 – LambdaMart – Cluster 3

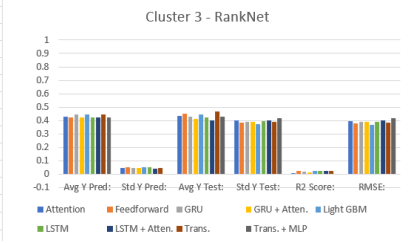


Figure 13c. Precision@10 – RankNet – Cluster 3

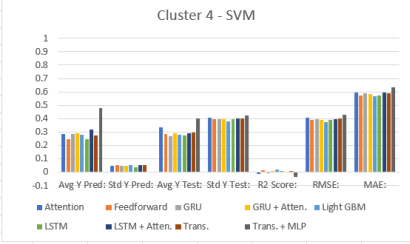


Figure 14a. Precision@10 – SVM – Cluster 4

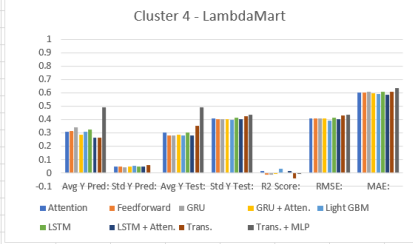


Figure 14b. Precision@10 – LambdaMart – Cluster 4

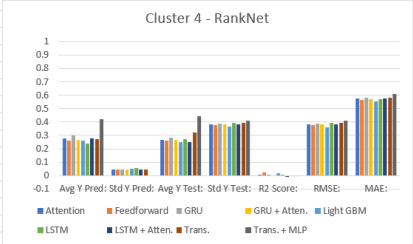


Figure 14c. Precision@10 – RankNet – Cluster 4

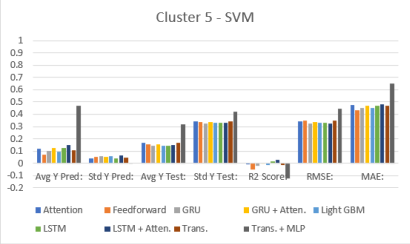


Figure 15a. Precision@10 – SVM – Cluster 5

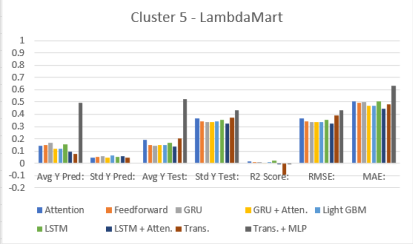


Figure 15b. Precision@10 – LambdaMart – Cluster 5

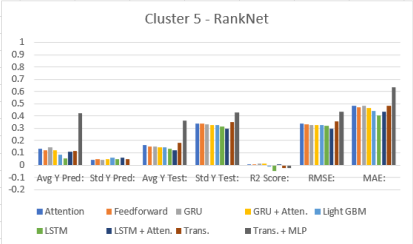


Figure 15c. Precision@10 – RankNet – Cluster 5

4.2 Normalized Distributed Cumulative Gain (NDCG)

4.2.1 NDCG@5

For NDCG@5, the LightGBM model on average outperforms the ANN models in terms of $R2$ score and $RMSE$ / MAE. However, when considering clustering, the feedforward network performs better for lower NDCG scores rather than higher performing *[Appendix 6.2.1, Table 7, Table 8, Table 9]* and has lower error than LightGBM in the lower cluster regions.

Table 7. RankingSVM - NDCG@5

NDCG@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0450	0.0399	0.0446	0.0448	0.0376	0.0447	0.0434	0.0491	0.0569
R2 Score:	0.2056	0.2965	0.2141	0.2098	0.3374	0.2113	0.2350	0.1333	-0.0044
RMSE:	0.2122	0.1997	0.2111	0.2117	0.1938	0.2114	0.2083	0.2217	0.2386
MAE:	0.4141	0.3997	0.4128	0.4140	0.3927	0.4130	0.4098	0.4210	0.4406

Table 8. LambdaMart - NDCG@5

NDCG@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0462	0.0404	0.0451	0.0453	0.0375	0.0442	0.0443	0.0580	0.0570
R2 Score:	0.1855	0.2876	0.2038	0.2007	0.3383	0.2208	0.2181	-0.0224	-0.0054
RMSE:	0.2149	0.2010	0.2125	0.2129	0.1937	0.2102	0.2105	0.2407	0.2387
MAE:	0.4160	0.4012	0.4148	0.4151	0.3925	0.4117	0.4125	0.4381	0.4408

Table 9. RankNet - NDCG@5

NDCG@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0452	0.0388	0.0450	0.0436	0.0375	0.0436	0.0440	0.0502	0.0570
R2 Score:	0.2031	0.3160	0.2068	0.2304	0.3379	0.2306	0.2247	0.1145	-0.0048
RMSE:	0.2125	0.1969	0.2121	0.2089	0.1937	0.2088	0.2096	0.2240	0.2387
MAE:	0.4146	0.3968	0.4138	0.4093	0.3930	0.4101	0.4108	0.4250	0.4407

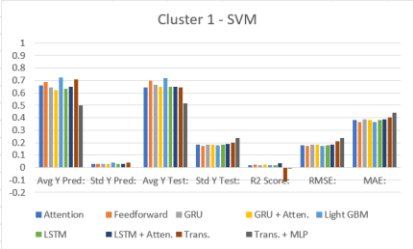


Figure 16a. NDCG@5 – SVM – Cluster 1

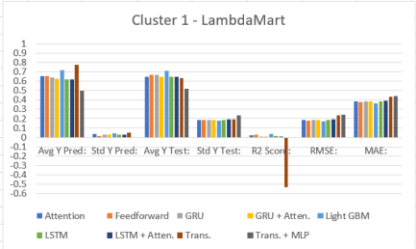


Figure 16b. NDCG@5 – LambdaMart – Cluster 1

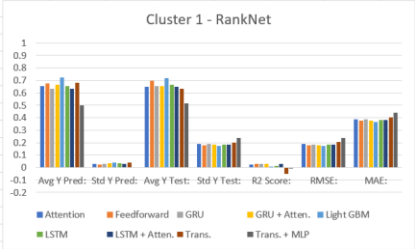


Figure 16c. NDCG@5 – RankNet – Cluster 1

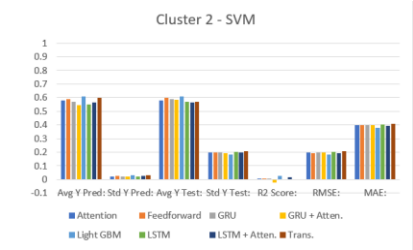


Figure 17a. NDCG@5 – SVM – Cluster 2

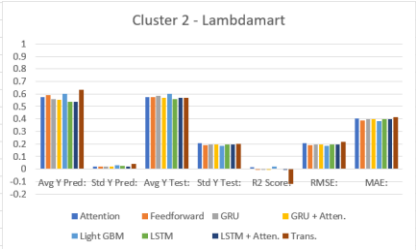


Figure 17b. NDCG@5 – LambdaMart – Cluster 2

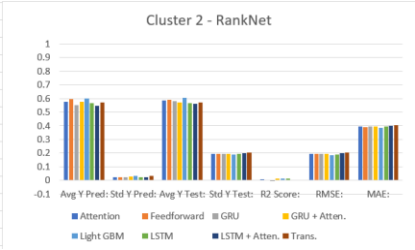


Figure 17c. NDCG@5 – RankNet – Cluster 2

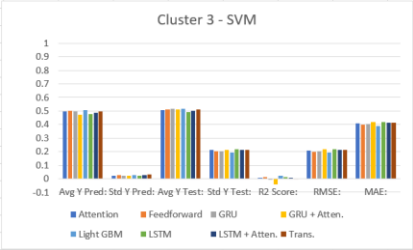


Figure 18a. NDCG@5 – SVM – Cluster 3

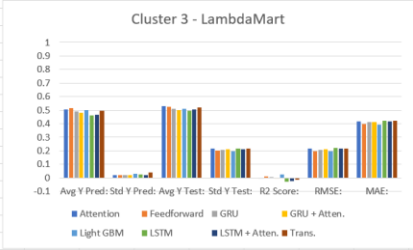


Figure 18b. NDCG@5 – LambdaMart – Cluster 3

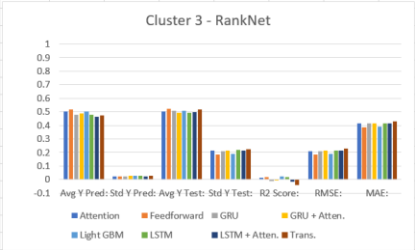


Figure 18c. NDCG@5 – RankNet – Cluster 3

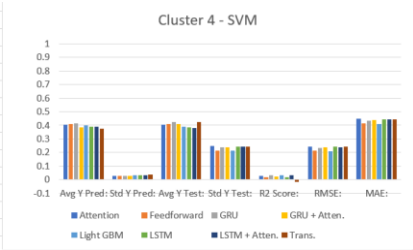


Figure 19a. NDCG@5 – SVM – Cluster 4

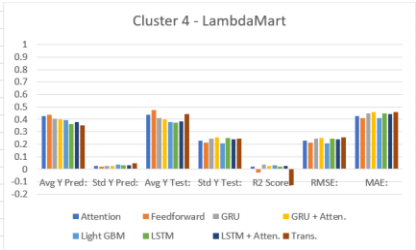


Figure 19b. NDCG@5 – LambdaMart – Cluster 4

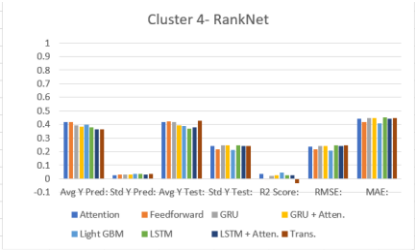


Figure 19c. NDCG@5 – RankNet – Cluster 4

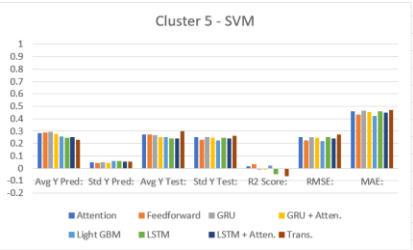


Figure 20a. NDCG@5 – SVM – Cluster 5

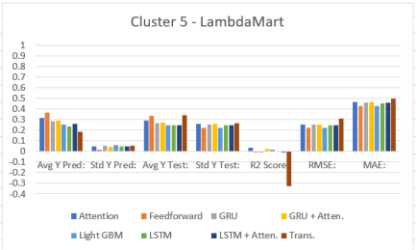


Figure 20b. NDCG@5 – LambdaMart – Cluster 5

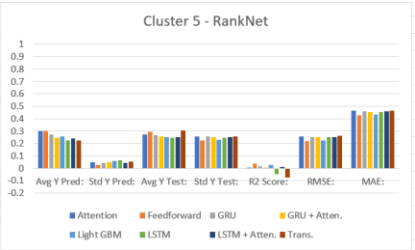


Figure 20c. NDCG@5 – RankNet – Cluster 5

4.2.2 NDCG@10

Considering the results for NDCG@10, the Light GBM model performs the best *[Table 10, Table 11, Table 12]*. Comparative to NDCG@5 *[Table 8, Table 9, Table 10]*, the Light GBM model performs performance between higher perceived and lower perceived performance is more uniform between clusters *[Appendix 6.2.2, Figure 21, Figure 22, Figure 23, Figure 24, Figure 25]*.

Table 20. RankingSVM - NDCG@10

NDCG@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0382	0.0279	0.0391	0.0382	0.0260	0.0380	0.0377	0.0432	0.0396
R2 Score:	0.2390	0.4433	0.2215	0.2394	0.4815	0.2418	0.2484	0.1387	0.2112
RMSE:	0.1954	0.1671	0.1976	0.1953	0.1613	0.1950	0.1942	0.2079	0.1989
MAE:	0.3964	0.3964	0.3985	0.3951	0.3521	0.3951	0.3943	0.4069	0.4007

Table 11. LambdaMart - NDCG@10

NDCG@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0377	0.0501	0.0335	0.0366	0.0260	0.0338	0.0381	0.0428	0.0473
R2 Score:	0.2482	0.0006	0.3313	0.2709	0.4825	0.3263	0.2397	0.1476	0.0573
RMSE:	0.1942	0.2239	0.1832	0.1912	0.1611	0.1838	0.1953	0.2068	0.2175
MAE:	0.3927	0.4254	0.3800	0.3907	0.3519	0.3815	0.3957	0.4055	0.4192

Table 12. RankNet - NDCG@10

NDCG@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0366	0.0270	0.0325	0.0363	0.0261	0.0332	0.0365	0.0442	0.0503
R2 Score:	0.2711	0.4610	0.3521	0.2757	0.4805	0.3381	0.2723	0.1184	-0.0022
RMSE:	0.1912	0.1644	0.1803	0.1906	0.1614	0.1822	0.1911	0.2103	0.2242
MAE:	0.3902	0.3569	0.3772	0.3900	0.3522	0.3797	0.3907	0.4068	0.4255

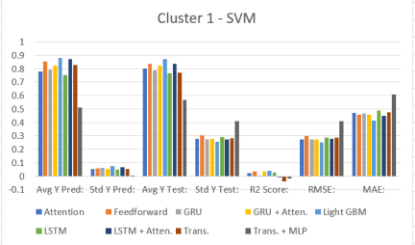


Figure 21a. NDCG@10 – SVM – Cluster 1

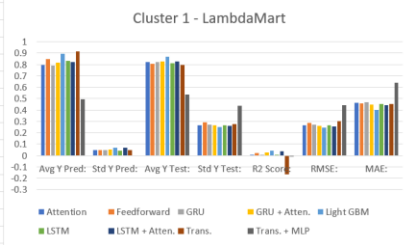


Figure 21b. NDCG@10 – LambdaMart – Cluster 1

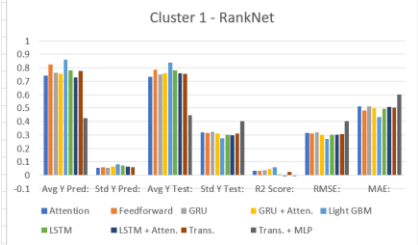


Figure 21c. NDCG@10 – RankNet – Cluster 1

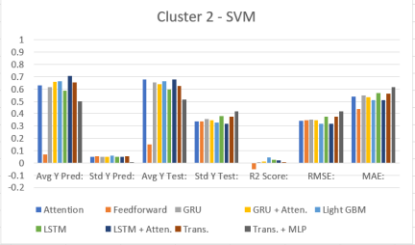


Figure 22a. NDCG@10 – SVM – Cluster 2

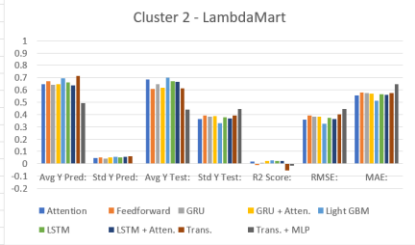


Figure 22b. NDCG@10 – LambdaMart – Cluster 2

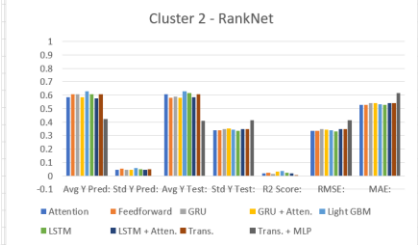


Figure 22c. NDCG@10 – RankNet – Cluster 2

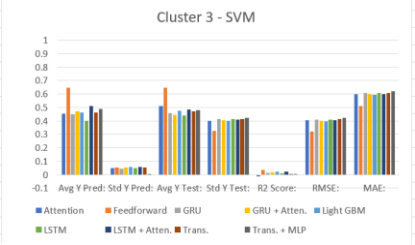


Figure 23a. NDCG@10 – SVM – Cluster 3

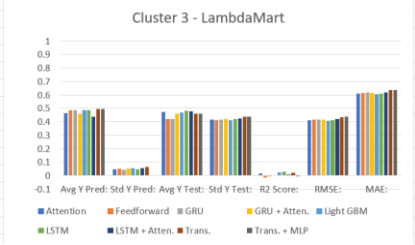


Figure 23b. NDCG@10 - LambdaMart – Cluster 3

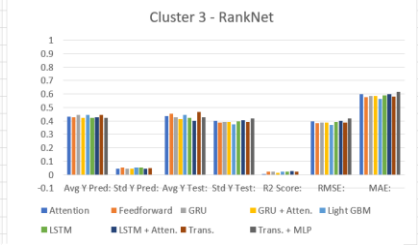


Figure 23c. NDCG@10 – RankNet – Cluster 3

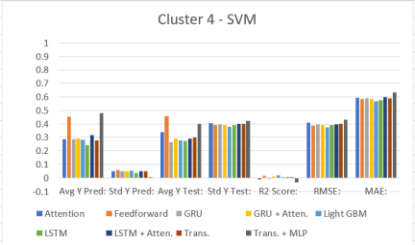


Figure 24a. NDCG@10 – SVM – Cluster 4

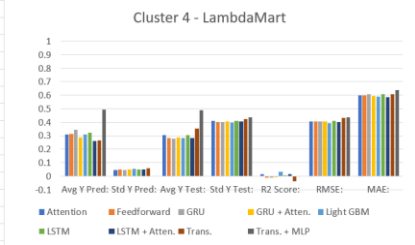


Figure 24b. NDCG@10 – LambdaMart – Cluster 4

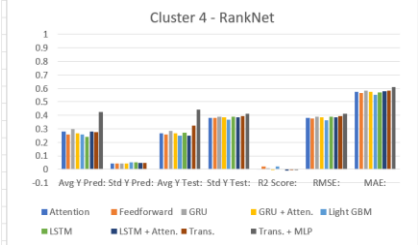


Figure 24c. NDCG@10 – RankNet – Cluster 4

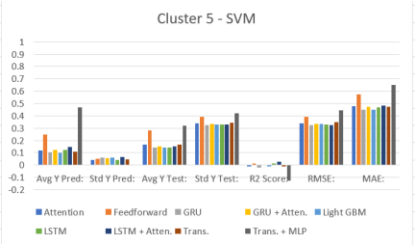


Figure 25a. NDCG@10 – SVM – Cluster 5

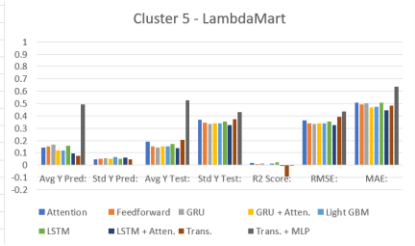


Figure 25b. NDCG@10 – LambdaMart – Cluster 5

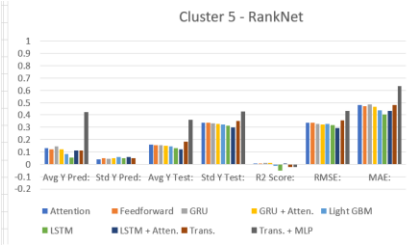


Figure 25c. NDCG@10 – RankNet – Cluster 5

4.3 Reciprocal Rank

4.3.1 RR@5

For RR@5, the LSTM + attention model outperforms the other models slightly however, the LightGBM model slightly outperforms the LSTM + attention model while utilizing the underlying RankingSVM case *[Table 13, Table 14, Table 15]*.

However, when considering clustering, the $R2$ scores are relatively weak on an order of $10e-3$ in some cases which suggests that the results are not significant *[Appendix 6.3.1, Figure 26, Figure 27, Figure 28, Figure 29, Figure 30]*. Noticeably, the overall error decreases as each model predicts the performance of the query document set to score worse.

Table 13. RankingSVM – RR@5

RR@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0642	0.0655	0.0631	0.0634	0.0626	0.0632	0.0628	0.0711	0.0657
R2 Score:	0.0623	0.0421	0.0778	0.0741	0.0844	0.0763	0.0828	-0.0395	0.0399
RMSE:	0.2533	0.2560	0.2512	0.2517	0.2503	0.2514	0.2505	0.2667	0.2563
MAE:	0.4364	0.4286	0.4340	0.4290	0.4286	0.4318	0.4280	0.4474	0.4382

Table 14. LambdaMart - RR@5

RR@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0644	0.0627	0.0632	0.0630	0.0626	0.0633	0.0626	0.0759	0.0632
R2 Score:	0.0590	0.0838	0.0761	0.0790	0.0855	0.0748	0.0857	-0.1093	0.0763
RMSE:	0.2537	0.2504	0.2514	0.2510	0.2502	0.2516	0.2501	0.2755	0.2514
MAE:	0.4312	0.4290	0.4293	0.4341	0.4286	0.4336	0.4280	0.4631	0.4306

Table 15. RankNet - RR@5

RR@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0645	0.0627	0.0642	0.0633	0.0628	0.0629	0.0626	0.0706	0.0656
R2 Score:	0.0575	0.0831	0.0615	0.0756	0.0823	0.0809	0.0848	-0.0318	0.0420
RMSE:	0.2539	0.2505	0.2534	0.2515	0.2506	0.2508	0.2502	0.2657	0.2560
MAE:	0.4362	0.4278	0.4381	0.4335	0.4286	0.4326	0.4318	0.4525	0.4383

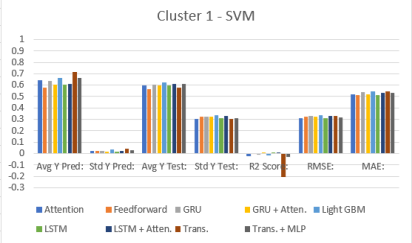


Figure 26a. RR@5 – SVM – Cluster 1

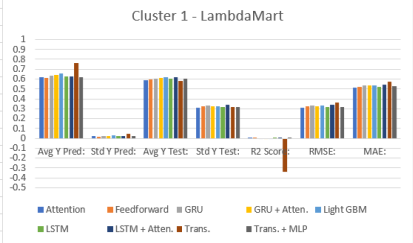


Figure 26b. RR@5 – LambdaMart – Cluster 1

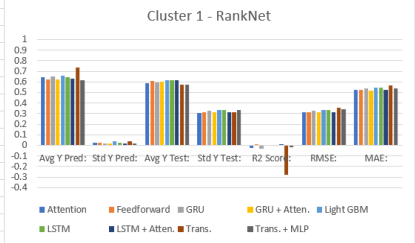


Figure 26c. RR@5 – RankNet – Cluster 1

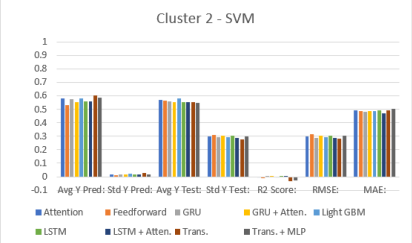


Figure 27a. RR@5 – SVM – Cluster 2

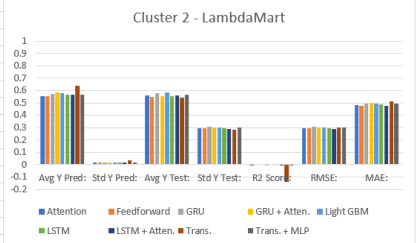


Figure 27b. RR@5 – LambdaMart – Cluster 2

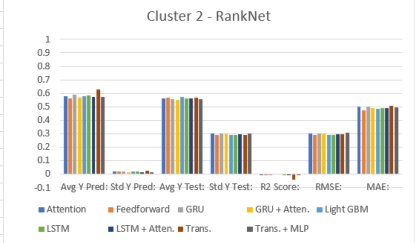


Figure 27c. RR@5 – RankNet – Cluster 2

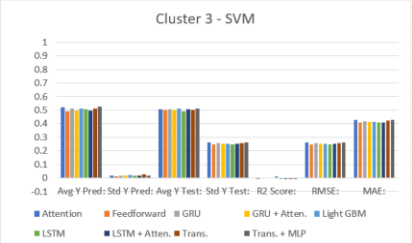


Figure 28a. RR@5 – SVM – Cluster 3

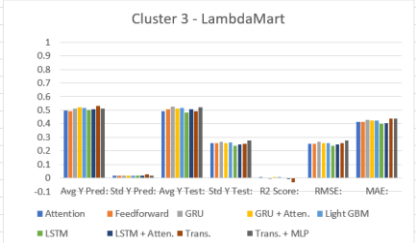


Figure 28b. RR@5 – LambdaMart – Cluster 3

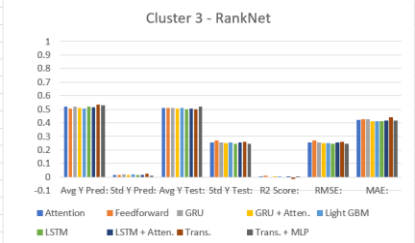


Figure 28c. RR@5 – RankNet – Cluster 3

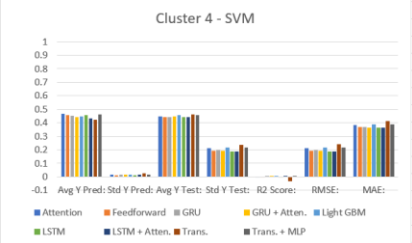


Figure 29a. RR@5 – SVM – Cluster 4

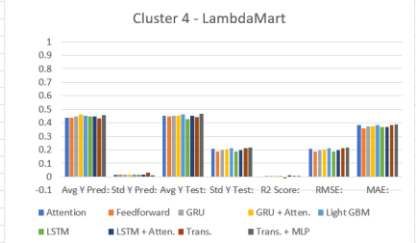


Figure 29a. RR@5 – LambdaMart – Cluster 4

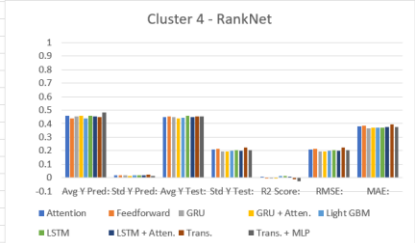


Figure 29c. RR@5 – RankNet – Cluster 4

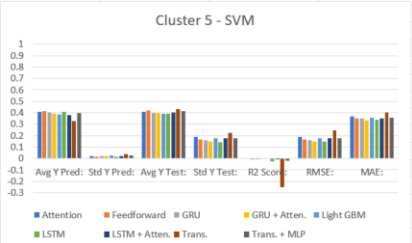


Figure 30a. RR@5 – SVM – Cluster 5

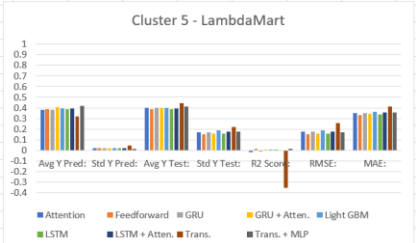


Figure 30b. RR@5 – LambdaMart – Cluster 5

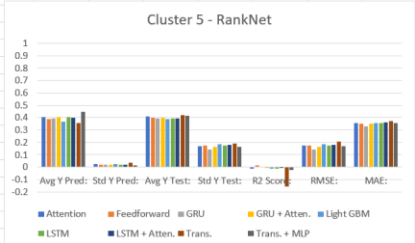


Figure 30c. RR@5 – RankNet – Cluster 5

4.3.1 RR@10

For RR@10, the Light GBM model in general has a stronger $R2$ score than the other models and an overall lower error *[Table 16, Table 17, Table 18]*. Noticeably, the average performance of each model is relatively similar here even considering the Transformer model which has not proven to be very effective across the different evaluation metrics often performing among the worst for this task.

When considering clusters, similar to RR@5, the model has low $R2$ scores across the different clusters *[Appendix 6.3.2, Figure 31, Figure 32, Figure 33, Figure 34, Figure 35]*. Furthermore, another similarity shows the overall error decreases as each model predicts the performance of the query document set to score worse.

Table 16. RankingSVM - RR@10

RR@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0430	0.0426	0.0411	0.0409	0.0397	0.0416	0.0412	0.0461	0.0414
R2 Score:	0.1188	0.1269	0.1579	0.1608	0.1846	0.1464	0.1547	0.0550	0.1517
RMSE:	0.2073	0.2063	0.2026	0.2023	0.1994	0.2040	0.2030	0.2146	0.2034
MAE:	0.3778	0.3742	0.3742	0.3736	0.3685	0.3729	0.3695	0.3928	0.3718

Table 17. LambdaMart - RR@10

RR@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0414	0.0418	0.0418	0.0412	0.0396	0.0412	0.0407	0.0496	0.0488
R2 Score:	0.1504	0.1426	0.1420	0.1557	0.1877	0.1559	0.1651	-0.0171	-0.0002
RMSE:	0.2035	0.2045	0.2045	0.2029	0.1990	0.2029	0.2017	0.2227	0.2208
MAE:	0.3722	0.3749	0.3715	0.3757	0.3681	0.3737	0.3713	0.3943	0.3986

Table 18. RankNet - RR@10

RR@10	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.0415	0.0424	0.0417	0.0410	0.0402	0.0413	0.0411	0.0454	0.0432
R2 Score:	0.1486	0.1295	0.1451	0.1585	0.1760	0.1529	0.1569	0.0679	0.1130
RMSE:	0.2037	0.2060	0.2041	0.2025	0.2004	0.2032	0.2027	0.2132	0.2080
MAE:	0.3710	0.3705	0.3712	0.3710	0.3687	0.3734	0.3712	0.3859	0.3735

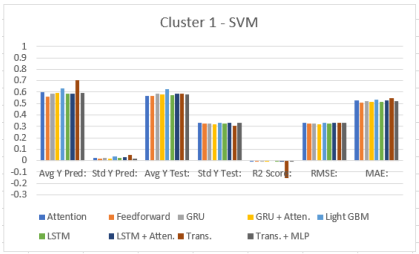


Figure 31a. RR@10 – SVM – Cluster 1

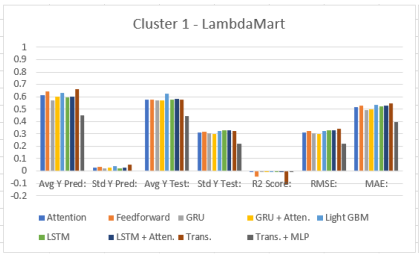


Figure 31b. RR@10 – LambdaMart – Cluster 1

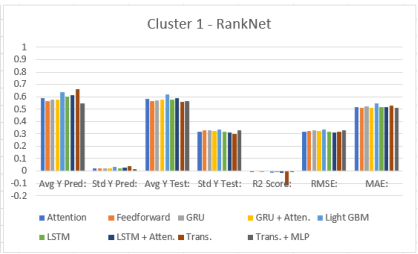


Figure 31c. RR@10 – RankNet – Cluster 1

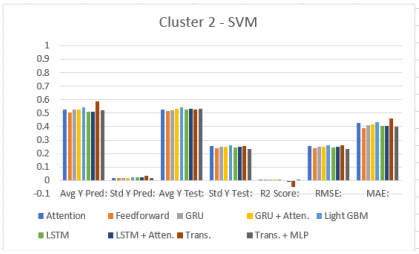


Figure 32a. RR@10 – SVM – Cluster 2

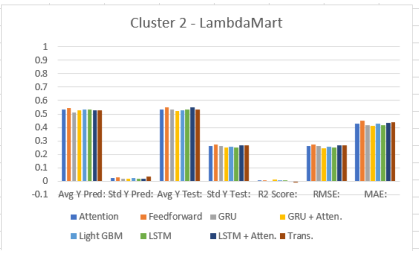


Figure 32b. RR@10 – LambdaMart – Cluster 2

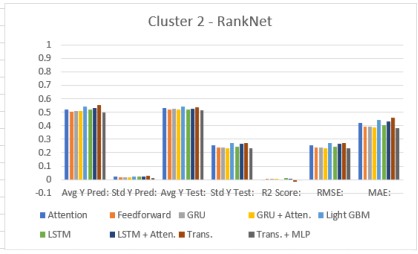


Figure 32c. RR@10 – RankNet – Cluster 2

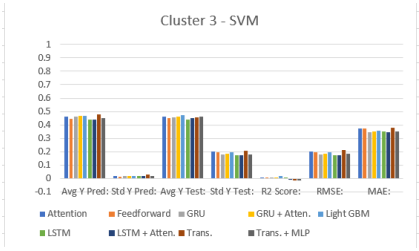


Figure 33a. RR@10 – SVM – Cluster 3

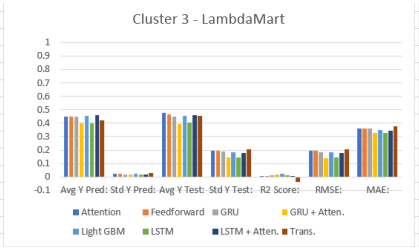


Figure 33b. RR@10 – LambdaMart – Cluster 3

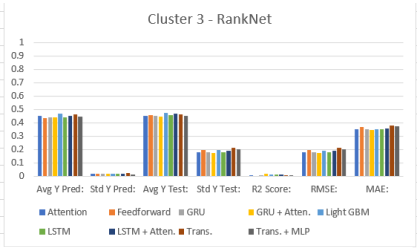


Figure 33c. RR@10 – RankNet – Cluster 3

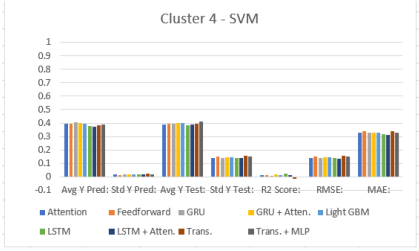


Figure 34a. RR@10 – SVM – Cluster 4

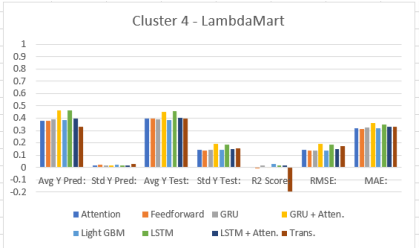


Figure 34b. RR@10 – LambdaMart – Cluster 4

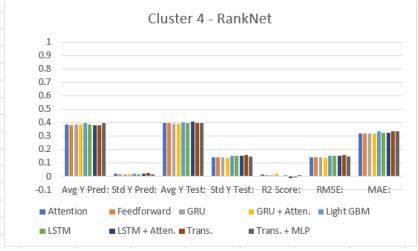


Figure 34c. RR@10 – RankNet – Cluster 4

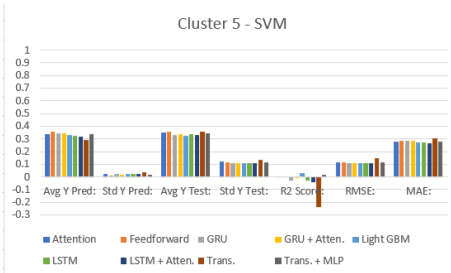


Figure 35a. RR@10 – SVM – Cluster 5

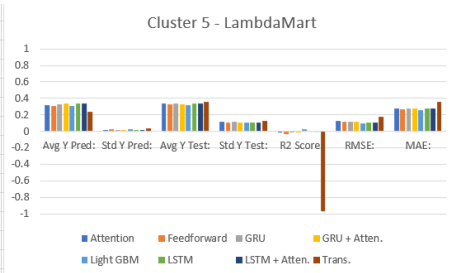


Figure 35b. RR@10 – LambdaMart – Cluster 5

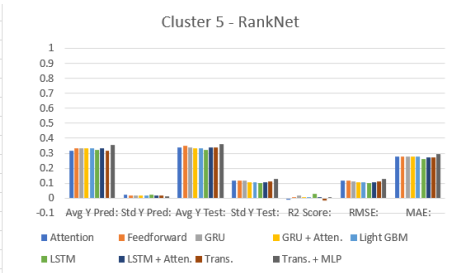


Figure 35c. RR@10 – RankNet – Cluster 5

4.4 Expected Reciprocal Rank (ERR)

4.4.1 ERR@5

For ERR@5, the $R2$ scores for this task are relatively similar for the Attention, LSTM, LSTM + Attention, GRU, GRU + Attention, and the Transformer Model + MLP (in the LambdaMart / RankNet cases) *[Table 19, Table 20, Table 21]*. However, in general, the LightGBM model on average has a stronger $R2$ and a lower error than the other models.

When considering on clusters, the overall $R2$ does not change much between clusters and is significantly worse which suggests that the model may have converged towards an average score for predicting the ERR *[Appendix 6.4.1, Figure 36, Figure 37, Figure 38, Figure 39]*.

Table 19. RankingSVM - ERR@5

ERR@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1779	0.2157	0.1762	0.1753	0.1728	0.1749	0.1753	0.1874	0.1757
R2 Score:	0.1755	0.0001	0.1834	0.1874	0.1990	0.1893	0.1875	0.1313	0.1855
RMSE:	0.4218	0.4644	0.4197	0.4187	0.4157	0.4182	0.4187	0.4329	0.4192
MAE:	0.5318	0.5749	0.5284	0.5289	0.5302	0.5281	0.5280	0.5437	0.5304

Table 20. LambdaMart - ERR@5

ERR@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1798	0.1740	0.1770	0.1746	0.1720	0.1748	0.1746	0.1811	0.2159
R2 Score:	0.1665	0.1934	0.1794	0.1906	0.2027	0.1895	0.1908	0.1606	-0.0008
RMSE:	0.4241	0.4171	0.4207	0.4179	0.4147	0.4181	0.4178	0.4255	0.4647
MAE:	0.5299	0.5280	0.5276	0.5281	0.5290	0.5260	0.5294	0.5342	0.5744

Table 21. RankNet - ERR@5

ERR@5	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1779	0.1734	0.1760	0.1754	0.1724	0.1749	0.1743	0.1823	0.2158
R2 Score:	0.1752	0.1961	0.1844	0.1870	0.2007	0.1893	0.1921	0.1551	-0.0005
RMSE:	0.4218	0.4165	0.4195	0.4188	0.4152	0.4182	0.4175	0.4269	0.4646
MAE:	0.5338	0.5269	0.5279	0.5287	0.5296	0.5274	0.5291	0.5372	0.5750

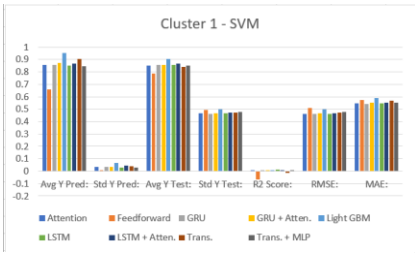


Figure 36a. ERR@5 – SVM – Cluster 1

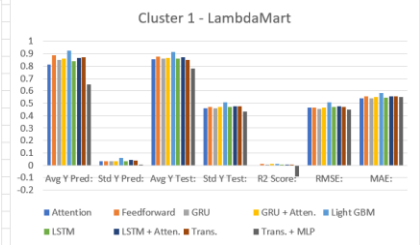


Figure 36b. ERR@5 – LambdaMart – Cluster 1

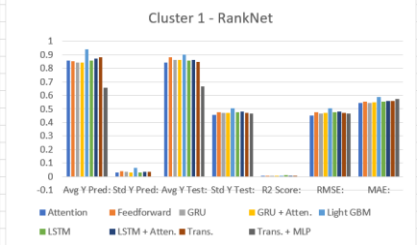


Figure 36c. ERR@5 – RankNet – Cluster 1

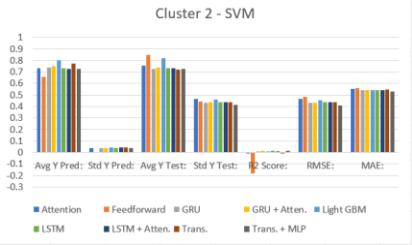


Figure 37a. ERR@5 – SVM – Cluster 2

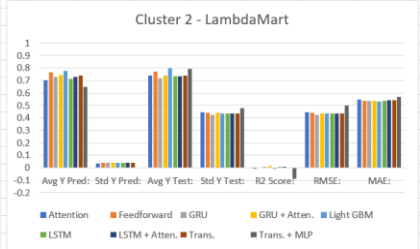


Figure 37b. ERR@5 – LambdaMart – Cluster 2

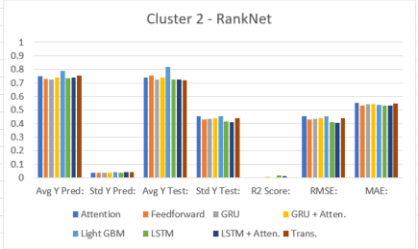


Figure 37c. ERR@5 – RankNet – Cluster 2

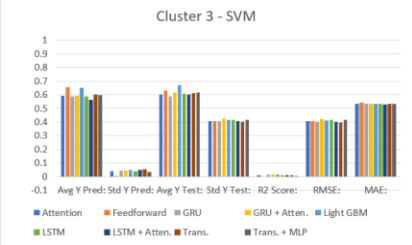


Figure 38a. ERR@5 – SVM – Cluster 3

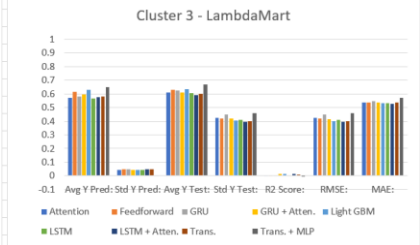


Figure 38b. ERR@5 – LambdaMart – Cluster 3

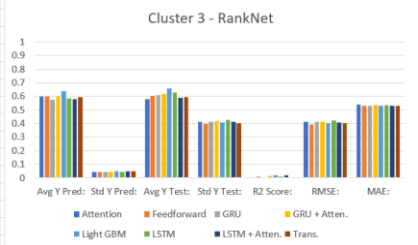


Figure 38c. ERR@5 – RankNet – Cluster 3

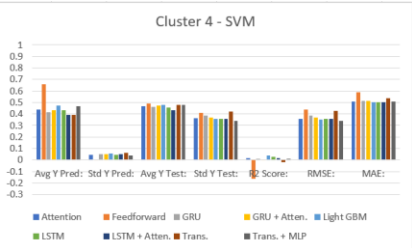


Figure 39a. ERR@5 – SVM – Cluster 4

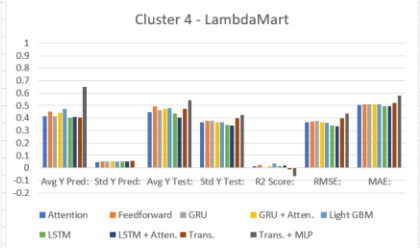


Figure 39b. ERR@5 – LambdaMart – Cluster 4

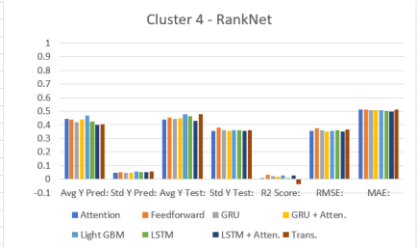


Figure 39c. ERR@5 – RankNet – Cluster 4

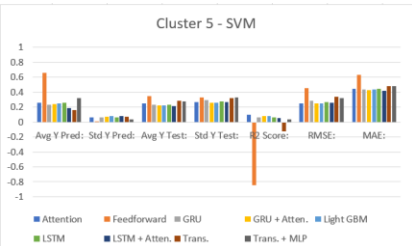


Figure 40a. ERR@5 – SVM – Cluster 5

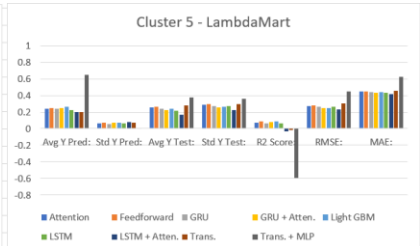


Figure 40b. ERR@5 – LambdaMart – Cluster 5

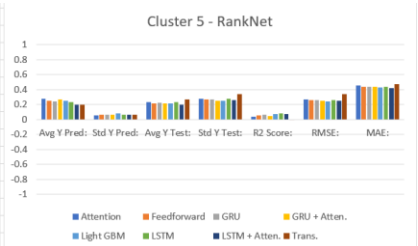


Figure 40c. ERR@5 – RankNet – Cluster 5

4.4.1 ERR@10

For ERR@10, the results are extremely similar to $ERR@5$ where the $R2$ scores for this task are relatively similar for the Attention, LSTM, LSTM + Attention, GRU, and GRU + Attention *[Table 22, Table 23, Table 24]*. Comparative to ERR@5, Feedforward performs similarly to the above-mentioned models whereas earlier, it had among the higher error rates. However, the LightGBM model on overall average has a stronger $R2$ and a lower error than the other models.

Table 22. RankingSVM - ERR@10

ERR@10	Attention	Feedforward	GRU	GRU + Atten.	LightGBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1720	0.1702	0.1716	0.1707	0.1680	0.1714	0.1710	0.1762	0.1756
R2 Score:	0.1748	0.1833	0.1767	0.1809	0.1939	0.1775	0.1794	0.1545	0.1575
RMSE:	0.4147	0.4126	0.4143	0.4132	0.4099	0.4141	0.4136	0.4198	0.4190
MAE:	0.5214	0.5179	0.5182	0.5197	0.5205	0.5204	0.5199	0.5281	0.5217

Table 23. LambdaMart - ERR@10

ERR@10	Attention	Feedforward	GRU	GRU + Atten.	LightGBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1724	0.1691	0.1710	0.1702	0.1686	0.1710	0.1706	0.1770	0.1745
R2 Score:	0.1730	0.1887	0.1796	0.1832	0.1912	0.1797	0.1817	0.1509	0.1627
RMSE:	0.4152	0.4112	0.4135	0.4126	0.4106	0.4135	0.4130	0.4207	0.4177
MAE:	0.5213	0.5152	0.5167	0.5192	0.5208	0.5207	0.5210	0.5290	0.5259

Table 24. RankNet - ERR@10

ERR@10	Attention	Feedforward	GRU	GRU + Atten.	LightGBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
MSE:	0.1734	0.1711	0.1715	0.1714	0.1672	0.1701	0.1702	0.1847	0.1760
R2 Score:	0.1681	0.1791	0.1774	0.1776	0.1978	0.1837	0.1832	0.1137	0.1558
RMSE:	0.4164	0.4136	0.4141	0.4140	0.4089	0.4125	0.4126	0.4298	0.4195
MAE:	0.5221	0.5160	0.5202	0.5181	0.5194	0.5235	0.5207	0.5363	0.5229

When considering the clusters, the $R2$ score performs stronger for perceived worse performing query documents *[Appendix 6.4.2, Figure 41, Figure 42, Figure 43, Figure 44, Figure 45]*. Furthermore, the standard deviation between the real query prediction score within clusters decreases as the model predicts a performing query document pair to be worse in performance.

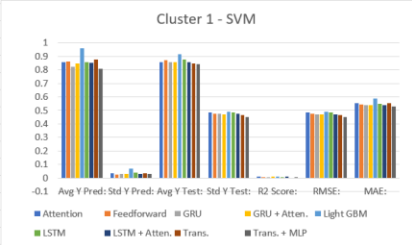


Figure 41a. ERR@10 – SVM – Cluster 1

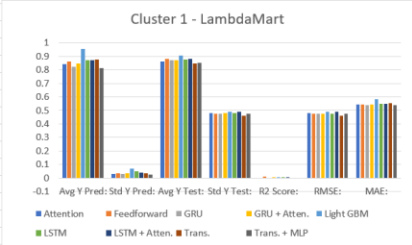


Figure 41b. ERR@5 – LambdaMart – Cluster 1

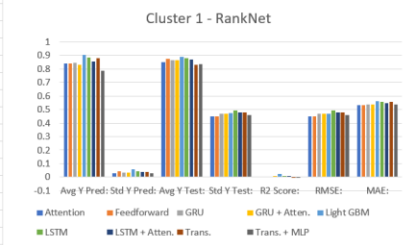


Figure 41c. ERR@5 – RankNet – Cluster 1

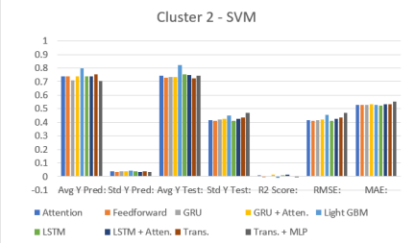


Figure 42a. ERR@10 – SVM – Cluster 2

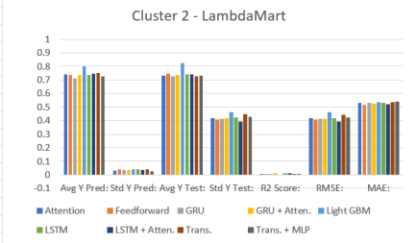


Figure 42b. ERR@5 – LambdaMart – Cluster 2

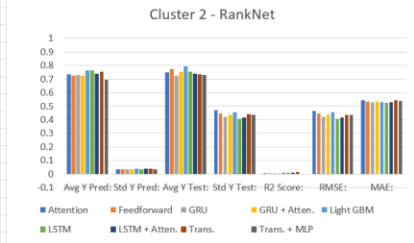


Figure 42c. ERR@5 – RankNet – Cluster 2

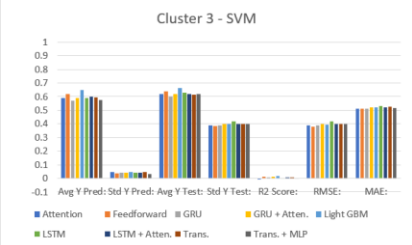


Figure 43a. ERR@10 – SVM – Cluster 3

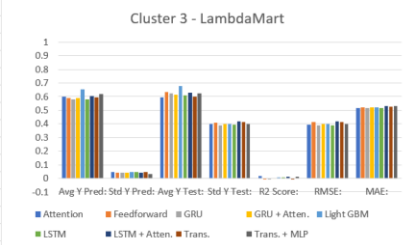


Figure 43b. ERR@5 – LambdaMart – Cluster 3

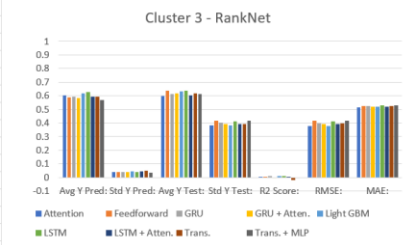


Figure 43c. ERR@5 – RankNet – Cluster 3

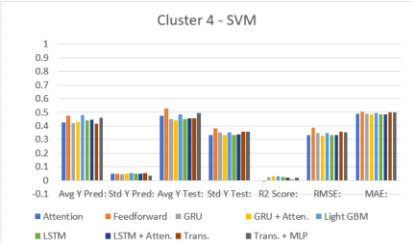


Figure 44a. ERR@10 – SVM – Cluster 4

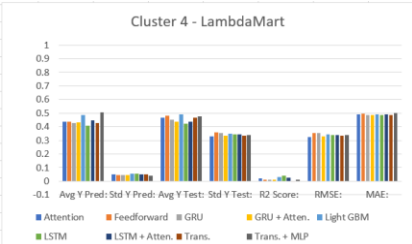


Figure 44b. ERR@5 – LambdaMart – Cluster 4

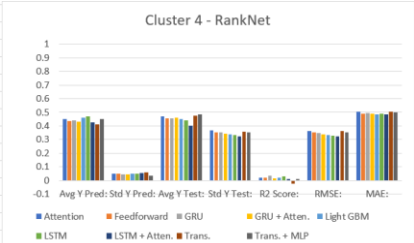


Figure 44c. ERR@5 – RankNet – Cluster 4

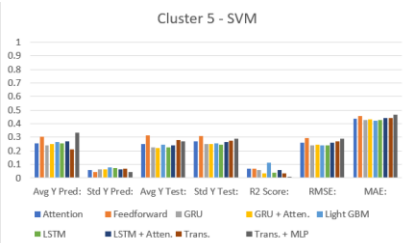


Figure 45a. ERR@10 – SVM – Cluster 5

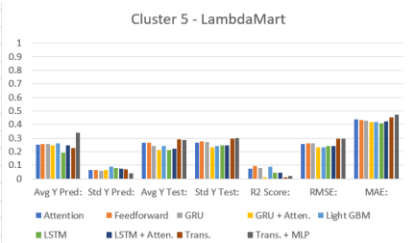


Figure 45b. ERR@5 – LambdaMart – Cluster 5

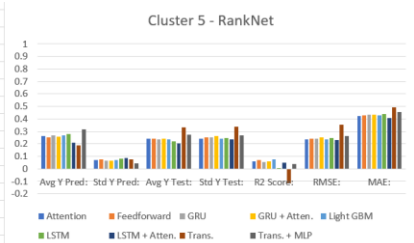


Figure 45c. ERR@5 – RankNet – Cluster 5

5 Conclusion and Future Work

In this thesis, we observe the feasibility of predicting standard LTR evaluation metrics given the list of features utilizing various ANN models and the gradient-boosted decision-tree LightGBM model. Similar to the conclusion that gradient-boosted decision-tree models outperform existing neural networks in LTR tasks [41], we observe that the tree model generally outperforms the ANNs in predicting the query performance prediction scores.

Furthermore, utilizing attention mechanisms and the state-of-the-art Transformer in this task, we show that these models underperform in the current task likely due to the lack of embedding guaranteeing distributional properties between sequences.

While the results suggested here are rough and not significant enough as a standalone prediction for evaluating query document query performance, it is worth exploring to see whether the predicted output performances for various models can help as additional feature inputs to enhance existing LTR models.

6 Appendix

6.1.1 Precision@5 Clusters

Cluster 1									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7686	0.6526	0.6975	0.7032	0.7884	0.6935	0.7100	0.7590	0.6440
Std Y Pred:	0.0546	0.0192	0.0577	0.0610	0.0679	0.0480	0.0535	0.0620	0.0341
Avg Y Test:	0.7300	0.6455	0.7074	0.7050	0.7747	0.6998	0.7243	0.6941	0.6570
Std Y Test:	0.2806	0.3201	0.2784	0.2889	0.2718	0.2875	0.2868	0.2956	0.3040
R2 Score:	0.0147	0.0202	0.0181	0.0422	0.0703	0.0200	0.0344	-0.0322	0.0176
RMSE:	0.2785	0.3169	0.2759	0.2827	0.2621	0.2846	0.2818	0.3003	0.3013
MAE:	0.4775	0.5115	0.4803	0.4833	0.4581	0.4872	0.4827	0.4930	0.4978
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7905	0.7916	0.7357	0.7725	0.8034	0.7390	0.7625	0.7872	0.7182
Std Y Pred:	0.0535	0.0504	0.0505	0.0565	0.0548	0.0592	0.0597	0.0558	0.0235
Avg Y Test:	0.7569	0.7682	0.7567	0.7545	0.7976	0.7375	0.7563	0.7503	0.7068
Std Y Test:	0.2765	0.2833	0.2762	0.2931	0.2593	0.2729	0.2709	0.2860	0.3015
R2 Score:	0.0034	0.0185	0.0074	0.0178	0.0421	-0.0262	0.0026	-0.0013	0.0102
RMSE:	0.2760	0.2806	0.2751	0.2904	0.2538	0.2764	0.2705	0.2862	0.2999
MAE:	0.4674	0.4692	0.4760	0.4814	0.4463	0.4783	0.4680	0.4761	0.4970
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6754	0.6426	0.6322	0.6793	0.7481	0.6202	0.6242	0.7087	0.5934
Std Y Pred:	0.0572	0.0550	0.0553	0.0573	0.0732	0.0549	0.0575	0.0624	0.0291
Avg Y Test:	0.6625	0.6527	0.6302	0.6691	0.7466	0.6373	0.6357	0.6457	0.5842
Std Y Test:	0.3016	0.3043	0.2968	0.3012	0.2903	0.3006	0.3005	0.3171	0.3028
R2 Score:	0.0126	0.0350	0.0451	0.0343	0.0410	-0.0015	0.0124	0.0152	0.0027
RMSE:	0.2997	0.2989	0.2900	0.2959	0.2843	0.3008	0.2986	0.3146	0.3024
MAE:	0.5013	0.4946	0.4858	0.4919	0.4810	0.4980	0.4948	0.5094	0.4950

Table 25. Precision@5 – Cluster 1

Cluster 2									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6133	0.5473	0.5400	0.5370	0.6085	0.5616	0.5663	0.5752	0.5372
Std Y Pred:	0.0495	0.0335	0.0487	0.0509	0.0513	0.0424	0.0472	0.0525	0.0342
Avg Y Test:	0.6034	0.4599	0.6056	0.5823	0.6194	0.5936	0.5796	0.5567	0.5269
Std Y Test:	0.3315	0.3997	0.3328	0.3366	0.3050	0.3385	0.3250	0.3500	0.3736
R2 Score:	0.0146	-0.0443	-0.0323	0.0162	0.0352	0.0019	0.0047	0.0207	0.0110
RMSE:	0.3291	0.4085	0.3381	0.3338	0.2995	0.3382	0.3243	0.3464	0.3716
MAE:	0.5212	0.6029	0.5323	0.5293	0.4947	0.5304	0.5193	0.5394	0.5666
Lambdamart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6165	0.6324	0.5901	0.6097	0.6489	0.5688	0.5971	0.6175	0.5912
Std Y Pred:	0.0547	0.0525	0.0447	0.0527	0.0497	0.0486	0.0499	0.0508	0.0388
Avg Y Test:	0.6510	0.6339	0.6589	0.6317	0.6606	0.6344	0.6345	0.6274	0.4949
Std Y Test:	0.3462	0.3425	0.3385	0.3468	0.3193	0.3570	0.3541	0.3511	0.4022
R2 Score:	-0.0023	0.0000	-0.0282	0.0320	0.0098	-0.0083	0.0110	0.0033	-0.0466
RMSE:	0.3466	0.3425	0.3433	0.3412	0.3178	0.3584	0.3522	0.3505	0.4115
MAE:	0.5429	0.5351	0.5415	0.5357	0.5129	0.5547	0.5485	0.5437	0.6015
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5287	0.5001	0.4849	0.5310	0.5625	0.4875	0.4776	0.5359	0.4824
Std Y Pred:	0.0431	0.0401	0.0428	0.0442	0.0500	0.0385	0.0426	0.0497	0.0344
Avg Y Test:	0.5310	0.5314	0.5253	0.5598	0.5717	0.5128	0.5026	0.5571	0.5152
Std Y Test:	0.3202	0.3396	0.3258	0.3161	0.3040	0.3325	0.3384	0.3143	0.3666
R2 Score:	0.0175	0.0011	-0.0096	-0.0027	0.0138	0.0010	0.0162	0.0155	0.0036
RMSE:	0.3174	0.3394	0.3274	0.3166	0.3019	0.3323	0.3357	0.3119	0.3660
MAE:	0.5076	0.5301	0.5167	0.5086	0.4901	0.5207	0.5261	0.5029	0.5587

Table 26. Precision@5 – Cluster 2

Cluster 3									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4277	0.4171	0.3620	0.3572	0.4239	0.4015	0.3757	0.3923	0.4060
Std Y Pred:	0.0544	0.0403	0.0461	0.0498	0.0530	0.0447	0.0520	0.0484	0.0379
Avg Y Test:	0.4371	0.3490	0.4036	0.3595	0.4171	0.3955	0.3855	0.3826	0.3663
Std Y Test:	0.3681	0.3691	0.3765	0.3792	0.3717	0.3773	0.3777	0.3799	0.3793
R2 Score:	0.0232	-0.0395	0.0092	0.0213	0.0422	0.0147	0.0216	0.0085	-0.0088
RMSE:	0.3638	0.3764	0.3748	0.3752	0.3638	0.3745	0.3736	0.3783	0.3810
MAE:	0.5596	0.5768	0.5691	0.5726	0.5611	0.5737	0.5697	0.5767	0.5804
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4202	0.4314	0.4224	0.4114	0.4512	0.4054	0.4122	0.4368	0.4400
Std Y Pred:	0.0548	0.0541	0.0472	0.0527	0.0516	0.0458	0.0500	0.0501	0.0448
Avg Y Test:	0.4236	0.4443	0.4671	0.3841	0.4104	0.4057	0.3790	0.4373	0.4076
Std Y Test:	0.3895	0.3834	0.3879	0.3927	0.3891	0.3979	0.3890	0.3960	0.3901
R2 Score:	0.0076	-0.0044	0.0232	0.0205	0.0186	0.0209	0.0066	0.0240	0.0016
RMSE:	0.3880	0.3842	0.3834	0.3886	0.3854	0.3937	0.3877	0.3912	0.3898
MAE:	0.5855	0.5790	0.5790	0.5901	0.5835	0.5919	0.5874	0.5891	0.5881
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3688	0.3605	0.3361	0.3708	0.3907	0.3493	0.3305	0.3634	0.3479
Std Y Pred:	0.0433	0.0404	0.0408	0.0452	0.0495	0.0377	0.0403	0.0488	0.0360
Avg Y Test:	0.3758	0.3621	0.3752	0.3906	0.3802	0.3663	0.3444	0.4103	0.4006
Std Y Test:	0.3600	0.3579	0.3656	0.3521	0.3471	0.3657	0.3582	0.3679	0.3743
R2 Score:	0.0145	0.0240	-0.0089	0.0204	0.0389	0.0018	0.0020	0.0056	-0.0199
RMSE:	0.3574	0.3535	0.3673	0.3485	0.3403	0.3654	0.3579	0.3669	0.3780
MAE:	0.5536	0.5502	0.5599	0.5398	0.5361	0.5609	0.5526	0.5574	0.5692

Table 27. Precision@5 – Cluster 3

Cluster 4									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2468	0.2771	0.2116	0.1899	0.2429	0.2513	0.2122	0.2304	0.2702
Std Y Pred:	0.0489	0.0362	0.0436	0.0475	0.0492	0.0415	0.0426	0.0446	0.0372
Avg Y Test:	0.2630	0.2920	0.2266	0.2076	0.2457	0.2465	0.2301	0.2410	0.2284
Std Y Test:	0.3571	0.3626	0.3506	0.3322	0.3475	0.3497	0.3395	0.3476	0.3422
R2 Score:	0.0185	-0.0092	-0.0003	0.0078	0.0130	0.0346	0.0179	0.0094	0.0056
RMSE:	0.3538	0.3642	0.3507	0.3309	0.3453	0.3436	0.3364	0.3459	0.3412
MAE:	0.5452	0.5569	0.5334	0.5106	0.5367	0.5370	0.5218	0.5337	0.5420
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2417	0.2520	0.2625	0.2419	0.2753	0.2457	0.2434	0.2626	0.2840
Std Y Pred:	0.0474	0.0468	0.0437	0.0451	0.0495	0.0464	0.0498	0.0499	0.0437
Avg Y Test:	0.2490	0.2824	0.2686	0.2227	0.2397	0.2068	0.1897	0.2669	0.2440
Std Y Test:	0.3603	0.3708	0.3759	0.3434	0.3547	0.3542	0.3411	0.3719	0.3680
R2 Score:	0.0308	0.0148	0.0289	0.0097	0.0085	-0.0137	-0.0252	0.0118	-0.0148
RMSE:	0.3548	0.3680	0.3705	0.3418	0.3531	0.3566	0.3454	0.3697	0.3707
MAE:	0.5471	0.5571	0.5627	0.5371	0.5540	0.5515	0.5423	0.5639	0.5711
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2276	0.2126	0.2014	0.2150	0.2239	0.2192	0.1971	0.2051	0.2289
Std Y Pred:	0.0399	0.0452	0.0355	0.0411	0.0461	0.0363	0.0375	0.0430	0.0294
Avg Y Test:	0.2231	0.2237	0.2385	0.2411	0.2264	0.2202	0.2182	0.2688	0.2424
Std Y Test:	0.3352	0.3350	0.3489	0.3422	0.3297	0.3359	0.3385	0.3489	0.3357
R2 Score:	0.0071	0.0142	0.0020	0.0028	0.0210	0.0238	0.0199	-0.0265	0.0096
RMSE:	0.3340	0.3326	0.3485	0.3417	0.3262	0.3319	0.3351	0.3534	0.3341
MAE:	0.5258	0.5185	0.5263	0.5283	0.5165	0.5210	0.5163	0.5321	0.5243

Table 28. Precision@5 – Cluster 4

Cluster 5									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.0877	0.1637	0.0611	0.0306	0.0802	0.1158	0.0787	0.0795	0.1507
Std Y Pred:	0.0403	0.0201	0.0464	0.0361	0.0538	0.0384	0.0352	0.0393	0.0232
Avg Y Test:	0.1349	0.1415	0.1147	0.1091	0.0977	0.1246	0.1125	0.1257	0.1237
Std Y Test:	0.2963	0.2949	0.2834	0.2775	0.2588	0.2918	0.2808	0.2917	0.2907
R2 Score:	-0.0146	0.0043	-0.0242	-0.0698	-0.0001	0.0172	-0.0037	-0.0143	0.0034
RMSE:	0.2985	0.2942	0.2868	0.2871	0.2588	0.2893	0.2813	0.2938	0.2902
MAE:	0.4253	0.4768	0.3883	0.3562	0.3889	0.4382	0.4017	0.4142	0.4648
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.0910	0.1047	0.1214	0.0973	0.1068	0.0825	0.0639	0.0978	0.1369
Std Y Pred:	0.0379	0.0400	0.0380	0.0394	0.0546	0.0517	0.0518	0.0446	0.0289
Avg Y Test:	0.1380	0.1257	0.1344	0.1104	0.1052	0.1030	0.0924	0.1308	0.1487
Std Y Test:	0.3056	0.2926	0.3001	0.2780	0.2733	0.2674	0.2516	0.2945	0.3162
R2 Score:	-0.0128	0.0127	0.0095	0.0119	0.0131	-0.0091	-0.0263	0.0014	0.0146
RMSE:	0.3076	0.2907	0.2987	0.2764	0.2715	0.2686	0.2549	0.2943	0.3138
MAE:	0.4345	0.4317	0.4521	0.4152	0.4180	0.3947	0.3669	0.4300	0.4722
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.0927	0.0471	0.0841	0.0808	0.0681	0.0974	0.0707	0.0649	0.1339
Std Y Pred:	0.0364	0.0431	0.0306	0.0353	0.0522	0.0329	0.0331	0.0336	0.0201
Avg Y Test:	0.1124	0.1053	0.1158	0.1145	0.0961	0.1109	0.1006	0.1158	0.1217
Std Y Test:	0.2768	0.2602	0.2771	0.2747	0.2561	0.2741	0.2637	0.2730	0.2858
R2 Score:	0.0030	-0.0571	0.0063	0.0004	-0.0089	0.0128	-0.0037	-0.0251	0.0101
RMSE:	0.2764	0.2676	0.2763	0.2747	0.2573	0.2724	0.2642	0.2764	0.2843
MAE:	0.4113	0.3647	0.4072	0.4023	0.3791	0.4138	0.3827	0.3907	0.4508

Table 29. Precision@5 – Cluster 5

6.1.2 Precision@10 Clusters

Cluster 1									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7805	0.8523	0.7939	0.8218	0.8782	0.7533	0.8694	0.8252	0.5123
Std Y Pred:	0.0512	0.0555	0.0623	0.0543	0.0735	0.0491	0.0649	0.0517	0.0058
Avg Y Test:	0.7989	0.8360	0.7881	0.8222	0.8689	0.7654	0.8352	0.7704	0.5656
Std Y Test:	0.2767	0.3048	0.2720	0.2764	0.2573	0.2893	0.2740	0.2812	0.4072
R2 Score:	0.0230	0.0366	0.0017	0.0368	0.0388	0.0253	-0.0095	-0.0383	-0.0158
RMSE:	0.2735	0.2992	0.2717	0.2713	0.2523	0.2856	0.2753	0.2865	0.4104
MAE:	0.4730	0.4559	0.4676	0.4582	0.4122	0.4866	0.4494	0.4744	0.6065
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7971	0.8474	0.7906	0.8158	0.8926	0.8312	0.8211	0.9149	0.4936
Std Y Pred:	0.0487	0.0483	0.0477	0.0517	0.0666	0.0444	0.0663	0.0463	0.0000
Avg Y Test:	0.8213	0.8066	0.8198	0.8254	0.8685	0.8096	0.8261	0.7960	0.5369
Std Y Test:	0.2665	0.2910	0.2732	0.2648	0.2522	0.2653	0.2607	0.2777	0.4378
R2 Score:	0.0078	0.0201	0.0024	0.0287	0.0438	0.0029	0.0353	-0.1674	-0.0098
RMSE:	0.2655	0.2880	0.2729	0.2610	0.2467	0.2649	0.2561	0.3001	0.4399
MAE:	0.4607	0.4567	0.4656	0.4500	0.4006	0.4541	0.4419	0.4533	0.6409
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7429	0.8266	0.7639	0.7557	0.8599	0.7794	0.7290	0.7770	0.4236
Std Y Pred:	0.0569	0.0609	0.0564	0.0627	0.0802	0.0731	0.0646	0.0595	0.0000
Avg Y Test:	0.7337	0.7849	0.7495	0.7575	0.8389	0.7789	0.7592	0.7564	0.4476
Std Y Test:	0.3182	0.3140	0.3253	0.3111	0.2774	0.3024	0.2994	0.3105	0.4043
R2 Score:	0.0331	0.0319	0.0386	0.0473	0.0593	0.0092	-0.0055	0.0249	-0.0035
RMSE:	0.3129	0.3089	0.3189	0.3037	0.2691	0.3010	0.3002	0.3066	0.4050
MAE:	0.5130	0.4824	0.5124	0.4990	0.4332	0.4944	0.5078	0.5023	0.6024

Table 30. Precision@10 – Cluster 1

Cluster 2									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6319	0.6464	0.6151	0.6600	0.6617	0.5887	0.7069	0.6530	0.4995
Std Y Pred:	0.0479	0.0538	0.0509	0.0505	0.0577	0.0518	0.0495	0.0535	0.0031
Avg Y Test:	0.6767	0.6464	0.6519	0.6386	0.6648	0.5946	0.6782	0.6244	0.5146
Std Y Test:	0.3394	0.3271	0.3546	0.3487	0.3272	0.3793	0.3205	0.3749	0.4189
R2 Score:	-0.0038	0.0352	0.0032	0.0124	0.0449	0.0259	0.0202	0.0068	-0.0011
RMSE:	0.3401	0.3213	0.3541	0.3465	0.3197	0.3744	0.3173	0.3736	0.4192
MAE:	0.5402	0.5123	0.5504	0.5360	0.5111	0.5693	0.5106	0.5642	0.6156
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6444	0.6681	0.6399	0.6452	0.6941	0.6601	0.6356	0.7150	0.4936
Std Y Pred:	0.0481	0.0529	0.0446	0.0536	0.0567	0.0504	0.0555	0.0604	0.0000
Avg Y Test:	0.6859	0.6084	0.6441	0.6198	0.7006	0.6722	0.6648	0.6122	0.4423
Std Y Test:	0.3629	0.3920	0.3850	0.3864	0.3293	0.3761	0.3675	0.3917	0.4434
R2 Score:	0.0164	-0.0117	0.0109	0.0254	0.0290	0.0222	0.0209	-0.0540	-0.0134
RMSE:	0.3600	0.3943	0.3828	0.3814	0.3245	0.3719	0.3636	0.4021	0.4463
MAE:	0.5566	0.5787	0.5760	0.5707	0.5145	0.5632	0.5594	0.5759	0.6473
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5866	0.6087	0.6081	0.5859	0.6304	0.6072	0.5758	0.6089	0.4236
Std Y Pred:	0.0441	0.0550	0.0451	0.0459	0.0585	0.0511	0.0434	0.0478	0.0000
Avg Y Test:	0.6056	0.5811	0.5921	0.5828	0.6308	0.6168	0.5855	0.6054	0.4081
Std Y Test:	0.3379	0.3386	0.3488	0.3510	0.3451	0.3338	0.3498	0.3487	0.4153
R2 Score:	0.0171	0.0245	0.0152	0.0315	0.0357	0.0225	0.0175	0.0064	-0.0014
RMSE:	0.3350	0.3344	0.3461	0.3454	0.3388	0.3300	0.3467	0.3476	0.4156
MAE:	0.5302	0.5267	0.5410	0.5411	0.5315	0.5292	0.5414	0.5436	0.6149

Table 31. Precision@10 – Cluster 2

Cluster 3									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4518	0.4525	0.4473	0.4709	0.4631	0.3987	0.5093	0.4620	0.4897
Std Y Pred:	0.0495	0.0590	0.0454	0.0522	0.0560	0.0501	0.0569	0.0538	0.0027
Avg Y Test:	0.5119	0.4556	0.4562	0.4454	0.4750	0.4393	0.4827	0.4703	0.4810
Std Y Test:	0.4006	0.3906	0.4134	0.4054	0.4019	0.4127	0.4083	0.4123	0.4235
R2 Score:	-0.0129	0.0143	0.0117	0.0202	0.0242	0.0073	0.0236	0.0009	0.0005
RMSE:	0.4031	0.3878	0.4110	0.4013	0.3970	0.4111	0.4034	0.4121	0.4234
MAE:	0.5985	0.5840	0.6084	0.5982	0.5941	0.6058	0.5988	0.6079	0.6220
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4641	0.4869	0.4869	0.4605	0.4884	0.4844	0.4384	0.4940	0.4936
Std Y Pred:	0.0472	0.0517	0.0423	0.0511	0.0550	0.0468	0.0563	0.0646	0.0000
Avg Y Test:	0.4738	0.4217	0.4201	0.4585	0.4664	0.4797	0.4762	0.4619	0.4610
Std Y Test:	0.4145	0.4125	0.4159	0.4181	0.4123	0.4187	0.4226	0.4362	0.4362
R2 Score:	0.0153	-0.0148	-0.0072	0.0025	0.0261	0.0272	0.0076	0.0193	-0.0056
RMSE:	0.4113	0.4156	0.4173	0.4176	0.4069	0.4130	0.4210	0.4319	0.4374
MAE:	0.6092	0.6131	0.6161	0.6137	0.6047	0.6113	0.6176	0.6336	0.6378
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4294	0.4254	0.4436	0.4244	0.4440	0.4229	0.4251	0.4448	0.4236
Std Y Pred:	0.0449	0.0515	0.0436	0.0465	0.0532	0.0532	0.0425	0.0489	0.0000
Avg Y Test:	0.4354	0.4519	0.4263	0.4145	0.4445	0.4239	0.4019	0.4681	0.4275
Std Y Test:	0.3993	0.3862	0.3914	0.3920	0.3744	0.3958	0.4037	0.3905	0.4173
R2 Score:	0.0054	0.0244	0.0206	0.0149	0.0246	0.0248	0.0251	0.0244	-0.0001
RMSE:	0.3982	0.3815	0.3874	0.3890	0.3698	0.3909	0.3986	0.3857	0.4173
MAE:	0.5964	0.5751	0.5867	0.5862	0.5628	0.5884	0.5983	0.5815	0.6178

Table 32. Precision@10 – Cluster 3

Cluster 4									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2861	0.2472	0.2878	0.2923	0.2810	0.2442	0.3188	0.2763	0.4802
Std Y Pred:	0.0477	0.0515	0.0489	0.0485	0.0521	0.0381	0.0513	0.0505	0.0030
Avg Y Test:	0.3374	0.2840	0.2662	0.2894	0.2781	0.2742	0.2914	0.2981	0.4002
Std Y Test:	0.4068	0.3938	0.3955	0.3933	0.3782	0.3937	0.3993	0.4024	0.4243
R2 Score:	-0.0135	0.0146	-0.0054	0.0089	0.0198	0.0068	0.0045	0.0075	-0.0345
RMSE:	0.4096	0.3909	0.3965	0.3916	0.3744	0.3924	0.3984	0.4009	0.4315
MAE:	0.5955	0.5721	0.5903	0.5855	0.5679	0.5755	0.5965	0.5905	0.6327
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3091	0.3135	0.3435	0.2864	0.3069	0.3237	0.2624	0.2637	0.4936
Std Y Pred:	0.0467	0.0486	0.0455	0.0502	0.0541	0.0481	0.0487	0.0591	0.0000
Avg Y Test:	0.3043	0.2815	0.2790	0.2873	0.2815	0.3045	0.2815	0.3515	0.4891
Std Y Test:	0.4097	0.4032	0.4032	0.4047	0.3980	0.4120	0.4036	0.4233	0.4363
R2 Score:	0.0166	-0.0120	-0.0119	-0.0080	0.0317	0.0061	0.0150	-0.0395	-0.0001
RMSE:	0.4063	0.4056	0.4056	0.4063	0.3917	0.4107	0.4005	0.4315	0.4364
MAE:	0.5996	0.6007	0.6077	0.5951	0.5884	0.6070	0.5844	0.6065	0.6367
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2785	0.2593	0.2975	0.2649	0.2584	0.2404	0.2785	0.2748	0.4236
Std Y Pred:	0.0442	0.0427	0.0437	0.0434	0.0523	0.0536	0.0452	0.0479	0.0000
Avg Y Test:	0.2658	0.2583	0.2846	0.2662	0.2512	0.2693	0.2497	0.3219	0.4438
Std Y Test:	0.3824	0.3798	0.3903	0.3848	0.3657	0.3914	0.3839	0.3954	0.4099
R2 Score:	0.0032	0.0206	0.0084	-0.0027	0.0200	0.0016	-0.0082	-0.0004	-0.0024
RMSE:	0.3818	0.3758	0.3887	0.3853	0.3620	0.3911	0.3854	0.3955	0.4104
MAE:	0.5742	0.5636	0.5824	0.5728	0.5532	0.5697	0.5792	0.5824	0.6078

Table 33. Precision@10 – Cluster 4

Cluster 5									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.1200	0.0681	0.1032	0.1243	0.0977	0.1239	0.1497	0.1087	0.4680
Std Y Pred:	0.0436	0.0541	0.0619	0.0545	0.0612	0.0426	0.0638	0.0469	0.0056
Avg Y Test:	0.1652	0.1527	0.1407	0.1529	0.1419	0.1438	0.1513	0.1682	0.3186
Std Y Test:	0.3403	0.3395	0.3233	0.3354	0.3308	0.3317	0.3318	0.3446	0.4200
R2 Score:	-0.0085	-0.0488	-0.0221	0.0061	-0.0111	0.0150	0.0299	-0.0127	-0.1233
RMSE:	0.3418	0.3477	0.3269	0.3344	0.3327	0.3292	0.3268	0.3467	0.4452
MAE:	0.4771	0.4359	0.4502	0.4721	0.4520	0.4686	0.4830	0.4718	0.6515
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.1435	0.1511	0.1690	0.1176	0.1192	0.1585	0.0967	0.0777	0.4936
Std Y Pred:	0.0467	0.0512	0.0578	0.0499	0.0662	0.0527	0.0590	0.0455	0.0000
Avg Y Test:	0.1917	0.1523	0.1437	0.1515	0.1506	0.1700	0.1400	0.2038	0.5251
Std Y Test:	0.3686	0.3422	0.3365	0.3392	0.3401	0.3558	0.3272	0.3737	0.4330
R2 Score:	0.0186	0.0085	0.0133	0.0019	0.0136	0.0210	-0.0004	-0.0934	-0.0053
RMSE:	0.3651	0.3407	0.3342	0.3389	0.3378	0.3520	0.3272	0.3907	0.4341
MAE:	0.5087	0.4946	0.5002	0.4702	0.4717	0.5072	0.4445	0.4844	0.6349
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.1319	0.1232	0.1449	0.1215	0.0829	0.0557	0.1105	0.1132	0.4236
Std Y Pred:	0.0414	0.0508	0.0455	0.0490	0.0613	0.0507	0.0612	0.0473	0.0000
Avg Y Test:	0.1616	0.1533	0.1535	0.1483	0.1462	0.1319	0.1207	0.1827	0.3616
Std Y Test:	0.3376	0.3357	0.3315	0.3264	0.3243	0.3117	0.2969	0.3516	0.4285
R2 Score:	0.0001	0.0035	0.0118	0.0116	-0.0120	-0.0495	0.0058	-0.0230	-0.0209
RMSE:	0.3376	0.3351	0.3296	0.3245	0.3263	0.3193	0.2960	0.3556	0.4329
MAE:	0.4828	0.4726	0.4838	0.4652	0.4402	0.4034	0.4355	0.4828	0.6370

Table 34. Precision@10 – Cluster 5

6.2.1 NDCG@5 Clusters

Cluster 1									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6619	0.6866	0.6444	0.6240	0.7259	0.6309	0.6486	0.7100	0.5018
Std Y Pred:	0.0312	0.0320	0.0301	0.0283	0.0408	0.0320	0.0289	0.0421	0.0000
Avg Y Test:	0.6453	0.6981	0.6643	0.6513	0.7188	0.6514	0.6510	0.6419	0.5176
Std Y Test:	0.1836	0.1738	0.1868	0.1842	0.1768	0.1825	0.1880	0.1978	0.2381
R2 Score:	0.0213	0.0217	0.0187	0.0223	0.0205	0.0176	0.0366	-0.1144	-0.0044
RMSE:	0.1816	0.1719	0.1850	0.1821	0.1750	0.1809	0.1845	0.2088	0.2386
MAE:	0.3807	0.3654	0.3846	0.3824	0.3684	0.3813	0.3849	0.4051	0.4406
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6567	0.6574	0.6396	0.6251	0.7187	0.6203	0.6168	0.7725	0.5001
Std Y Pred:	0.0325	0.0147	0.0317	0.0273	0.0422	0.0287	0.0274	0.0500	0.0000
Avg Y Test:	0.6499	0.6675	0.6660	0.6499	0.7128	0.6478	0.6509	0.6354	0.5176
Std Y Test:	0.1883	0.1824	0.1848	0.1846	0.1745	0.1852	0.1893	0.1899	0.2381
R2 Score:	0.0223	0.0273	0.0061	0.0077	0.0363	0.0121	0.0061	-0.5332	-0.0054
RMSE:	0.1862	0.1799	0.1842	0.1839	0.1713	0.1841	0.1888	0.2352	0.2387
MAE:	0.3843	0.3772	0.3840	0.3852	0.3635	0.3852	0.3912	0.4329	0.4408
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6542	0.6742	0.6329	0.6645	0.7214	0.6531	0.6310	0.6795	0.5012
Std Y Pred:	0.0306	0.0222	0.0314	0.0328	0.0410	0.0337	0.0307	0.0427	0.0000
Avg Y Test:	0.6471	0.6944	0.6538	0.6560	0.7170	0.6638	0.6473	0.6335	0.5176
Std Y Test:	0.1907	0.1806	0.1872	0.1820	0.1751	0.1841	0.1863	0.1992	0.2381
R2 Score:	0.0225	0.0282	0.0277	0.0318	0.0077	0.0158	0.0310	-0.0499	-0.0048
RMSE:	0.1886	0.1781	0.1846	0.1791	0.1744	0.1826	0.1834	0.2041	0.2387
MAE:	0.3881	0.3744	0.3853	0.3755	0.3674	0.3810	0.3835	0.4019	0.4407

Table 35. NDCG@5 – Cluster 1

Cluster 2								
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.5776	0.5912	0.5681	0.5465	0.6094	0.5502	0.5666	0.5995
Std Y Pred:	0.0236	0.0266	0.0203	0.0216	0.0316	0.0219	0.0239	0.0305
Avg Y Test:	0.5815	0.5967	0.5909	0.5820	0.6081	0.5715	0.5639	0.5706
Std Y Test:	0.1990	0.1953	0.1972	0.1942	0.1847	0.2005	0.1967	0.2067
R2 Score:	0.0055	0.0087	0.0022	-0.0205	0.0260	-0.0024	0.0181	-0.0022
RMSE:	0.1985	0.1944	0.1970	0.1962	0.1823	0.2008	0.1949	0.2069
MAE:	0.4004	0.3957	0.3990	0.3992	0.3800	0.4027	0.3954	0.4059
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.5750	0.5919	0.5609	0.5509	0.6016	0.5394	0.5393	0.6323
Std Y Pred:	0.0213	0.0210	0.0217	0.0203	0.0314	0.0230	0.0218	0.0402
Avg Y Test:	0.5762	0.5730	0.5851	0.5718	0.6007	0.5575	0.5682	0.5671
Std Y Test:	0.2054	0.1890	0.1956	0.1958	0.1862	0.1951	0.1950	0.2031
R2 Score:	0.0123	-0.0075	-0.0063	-0.0015	0.0216	0.0028	-0.0050	-0.1213
RMSE:	0.2041	0.1897	0.1962	0.1959	0.1841	0.1948	0.1955	0.2150
MAE:	0.4039	0.3882	0.3984	0.3964	0.3832	0.3963	0.3965	0.4134
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.5755	0.5944	0.5537	0.5756	0.6023	0.5654	0.5459	0.5725
Std Y Pred:	0.0217	0.0228	0.0219	0.0252	0.0306	0.0243	0.0236	0.0305
Avg Y Test:	0.5856	0.5917	0.5794	0.5708	0.6074	0.5668	0.5617	0.5688
Std Y Test:	0.1961	0.1922	0.1929	0.1946	0.1868	0.1919	0.1975	0.2020
R2 Score:	0.0080	0.0043	-0.0051	0.0126	0.0123	0.0149	0.0047	0.0008
RMSE:	0.1953	0.1918	0.1934	0.1934	0.1856	0.1905	0.1970	0.2020
MAE:	0.3966	0.3889	0.3955	0.3957	0.3840	0.3923	0.3989	0.4041

Table 36. NDCG@5 – Cluster 2

Cluster 3								
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.4961	0.5005	0.4975	0.4720	0.5073	0.4771	0.4855	0.4953
Std Y Pred:	0.0245	0.0262	0.0212	0.0234	0.0288	0.0229	0.0254	0.0311
Avg Y Test:	0.5070	0.5096	0.5169	0.5101	0.5167	0.4903	0.5012	0.5122
Std Y Test:	0.2113	0.2018	0.2039	0.2130	0.1954	0.2180	0.2140	0.2128
R2 Score:	0.0062	0.0130	-0.0061	-0.0390	0.0245	0.0126	0.0082	0.0022
RMSE:	0.2106	0.2005	0.2045	0.2171	0.1930	0.2166	0.2131	0.2126
MAE:	0.4110	0.4004	0.4064	0.4184	0.3917	0.4175	0.4151	0.4149
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.5050	0.5133	0.4888	0.4831	0.4996	0.4591	0.4657	0.4968
Std Y Pred:	0.0213	0.0222	0.0221	0.0209	0.0290	0.0255	0.0225	0.0409
Avg Y Test:	0.5317	0.5232	0.5104	0.4989	0.5101	0.4948	0.5033	0.5221
Std Y Test:	0.2152	0.1996	0.2086	0.2121	0.1978	0.2182	0.2123	0.2166
R2 Score:	-0.0041	0.0116	0.0012	-0.0021	0.0262	-0.0264	-0.0255	-0.0118
RMSE:	0.2156	0.1984	0.2085	0.2123	0.1952	0.2210	0.2150	0.2179
MAE:	0.4169	0.4000	0.4110	0.4145	0.3945	0.4231	0.4162	0.4199
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.5014	0.5173	0.4792	0.4867	0.5022	0.4809	0.4646	0.4718
Std Y Pred:	0.0227	0.0233	0.0230	0.0275	0.0281	0.0266	0.0257	0.0294
Avg Y Test:	0.5042	0.5248	0.5076	0.4940	0.5065	0.4939	0.4968	0.5184
Std Y Test:	0.2131	0.1876	0.2106	0.2140	0.1917	0.2175	0.2134	0.2243
R2 Score:	0.0117	0.0186	-0.0100	-0.0013	0.0230	0.0167	-0.0175	-0.0409
RMSE:	0.2119	0.1859	0.2116	0.2141	0.1895	0.2157	0.2153	0.2288
MAE:	0.4138	0.3850	0.4129	0.4145	0.3894	0.4177	0.4163	0.4317

Table 37. NDCG@5 – Cluster 3

Cluster 4								
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.4031	0.4083	0.4149	0.3873	0.4019	0.3880	0.3875	0.3761
Std Y Pred:	0.0296	0.0299	0.0275	0.0269	0.0349	0.0316	0.0326	0.0372
Avg Y Test:	0.4051	0.4075	0.4241	0.4076	0.3877	0.3831	0.3786	0.4250
Std Y Test:	0.2477	0.2154	0.2387	0.2402	0.2112	0.2438	0.2432	0.2419
R2 Score:	0.0291	0.0184	0.0305	0.0237	0.0329	0.0192	0.0302	-0.0176
RMSE:	0.2441	0.2134	0.2350	0.2373	0.2077	0.2414	0.2395	0.2440
MAE:	0.4508	0.4141	0.4358	0.4411	0.4102	0.4460	0.4437	0.4454
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.4265	0.4362	0.4048	0.4022	0.3936	0.3619	0.3792	0.3517
Std Y Pred:	0.0265	0.0213	0.0276	0.0270	0.0348	0.0308	0.0295	0.0455
Avg Y Test:	0.4372	0.4722	0.4116	0.4000	0.3780	0.3732	0.3818	0.4414
Std Y Test:	0.2295	0.2109	0.2472	0.2543	0.2104	0.2482	0.2424	0.2427
R2 Score:	0.0231	-0.0294	0.0363	0.0289	0.0318	0.0213	0.0259	-0.1270
RMSE:	0.2268	0.2140	0.2426	0.2506	0.2070	0.2455	0.2392	0.2576
MAE:	0.4286	0.4120	0.4465	0.4578	0.4093	0.4483	0.4418	0.4594
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.4170	0.4192	0.3908	0.3826	0.3993	0.3803	0.3656	0.3648
Std Y Pred:	0.0277	0.0309	0.0290	0.0332	0.0341	0.0337	0.0315	0.0349
Avg Y Test:	0.4165	0.4249	0.4157	0.3921	0.3879	0.3701	0.3786	0.4254
Std Y Test:	0.2433	0.2172	0.2454	0.2449	0.2122	0.2482	0.2435	0.2417
R2 Score:	0.0356	-0.0020	0.0214	0.0250	0.0478	0.0270	0.0282	-0.0345
RMSE:	0.2390	0.2174	0.2427	0.2419	0.2071	0.2448	0.2401	0.2458
MAE:	0.4427	0.4193	0.4457	0.4462	0.4092	0.4510	0.4429	0.4474

Table 38. NDCG@5 – Cluster 4

Cluster 5								
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.2826	0.2874	0.2959	0.2763	0.2565	0.2478	0.2507	0.2301
Std Y Pred:	0.0475	0.0458	0.0479	0.0428	0.0606	0.0568	0.0520	0.0546
Avg Y Test:	0.2715	0.2732	0.2659	0.2502	0.2500	0.2411	0.2429	0.2977
Std Y Test:	0.2539	0.2277	0.2503	0.2439	0.2229	0.2476	0.2419	0.2631
R2 Score:	0.0188	0.0345	-0.0097	-0.0079	0.0196	-0.0450	-0.0028	-0.0648
RMSE:	0.2515	0.2237	0.2515	0.2449	0.2207	0.2531	0.2422	0.2715
MAE:	0.4593	0.4324	0.4625	0.4559	0.4250	0.4594	0.4507	0.4680
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.3129	0.3683	0.2866	0.2930	0.2512	0.2352	0.2575	0.1858
Std Y Pred:	0.0481	0.0147	0.0511	0.0411	0.0585	0.0452	0.0455	0.0547
Avg Y Test:	0.2906	0.3327	0.2654	0.2733	0.2441	0.2461	0.2459	0.3421
Std Y Test:	0.2568	0.2233	0.2502	0.2593	0.2245	0.2448	0.2467	0.2679
R2 Score:	0.0357	-0.0010	-0.0102	0.0230	0.0167	0.0018	-0.0035	-0.3274
RMSE:	0.2522	0.2234	0.2515	0.2563	0.2226	0.2445	0.2471	0.3087
MAE:	0.4624	0.4296	0.4619	0.4649	0.4257	0.4515	0.4585	0.4951
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.
Avg Y Pred:	0.2988	0.3014	0.2728	0.2439	0.2564	0.2272	0.2382	0.2250
Std Y Pred:	0.0460	0.0294	0.0456	0.0509	0.0607	0.0654	0.0452	0.0514
Avg Y Test:	0.2733	0.2939	0.2657	0.2562	0.2510	0.2440	0.2515	0.3032
Std Y Test:	0.2553	0.2248	0.2551	0.2504	0.2301	0.2457	0.2538	0.2549
R2 Score:	0.0043	0.0372	0.0159	0.0030	0.0278	-0.0452	0.0093	-0.0743
RMSE:	0.2548	0.2206	0.2531	0.2500	0.2269	0.2512	0.2526	0.2642
MAE:	0.4671	0.4284	0.4618	0.4530	0.4327	0.4526	0.4588	0.4647

Table 39. NDCG@5 – Cluster 5

6.2.2 NDCG@10 Clusters

Cluster 1									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7805	0.8523	0.7939	0.8218	0.8782	0.7533	0.8694	0.8252	0.5123
Std Y Pred:	0.0512	0.0555	0.0623	0.0543	0.0735	0.0491	0.0649	0.0517	0.0058
Avg Y Test:	0.7989	0.8360	0.7881	0.8222	0.8689	0.7654	0.8352	0.7704	0.5656
Std Y Test:	0.2767	0.3048	0.2720	0.2764	0.2573	0.2893	0.2740	0.2812	0.4072
R2 Score:	0.0230	0.0366	0.0017	0.0368	0.0388	0.0253	-0.0095	-0.0383	-0.0158
RMSE:	0.2735	0.2992	0.2717	0.2713	0.2523	0.2856	0.2753	0.2865	0.4104
MAE:	0.4730	0.4559	0.4676	0.4582	0.4122	0.4866	0.4494	0.4744	0.6065
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7971	0.8474	0.7906	0.8158	0.8926	0.8312	0.8211	0.9149	0.4936
Std Y Pred:	0.0487	0.0483	0.0477	0.0517	0.0666	0.0444	0.0663	0.0463	0.0000
Avg Y Test:	0.8213	0.8066	0.8198	0.8254	0.8685	0.8096	0.8261	0.7960	0.5369
Std Y Test:	0.2665	0.2910	0.2732	0.2648	0.2522	0.2653	0.2607	0.2777	0.4378
R2 Score:	0.0078	0.0201	0.0024	0.0287	0.0438	0.0029	0.0353	-0.1674	-0.0098
RMSE:	0.2655	0.2880	0.2729	0.2610	0.2467	0.2649	0.2561	0.3001	0.4399
MAE:	0.4607	0.4567	0.4656	0.4500	0.4006	0.4541	0.4419	0.4533	0.6409
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7429	0.8266	0.7639	0.7557	0.8599	0.7794	0.7290	0.7770	0.4236
Std Y Pred:	0.0569	0.0609	0.0564	0.0627	0.0802	0.0731	0.0646	0.0595	0.0000
Avg Y Test:	0.7337	0.7849	0.7495	0.7575	0.8389	0.7789	0.7592	0.7564	0.4476
Std Y Test:	0.3182	0.3140	0.3253	0.3111	0.2774	0.3024	0.2994	0.3105	0.4043
R2 Score:	0.0331	0.0319	0.0386	0.0473	0.0593	0.0092	-0.0055	0.0249	-0.0035
RMSE:	0.3129	0.3089	0.3189	0.3037	0.2691	0.3010	0.3002	0.3066	0.4050
MAE:	0.5130	0.4824	0.5124	0.4990	0.4332	0.4944	0.5078	0.5023	0.6024

Table 40. NDCG@10 – Cluster 1

Cluster 2									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6319	0.0681	0.6151	0.6600	0.6617	0.5887	0.7069	0.6530	0.4995
Std Y Pred:	0.0479	0.0541	0.0509	0.0505	0.0577	0.0518	0.0495	0.0535	0.0031
Avg Y Test:	0.6767	0.1527	0.6519	0.6386	0.6648	0.5946	0.6782	0.6244	0.5146
Std Y Test:	0.3394	0.3395	0.3546	0.3487	0.3272	0.3793	0.3205	0.3749	0.4189
R2 Score:	-0.0038	-0.0488	0.0032	0.0124	0.0449	0.0259	0.0202	0.0068	-0.0011
RMSE:	0.3401	0.3477	0.3541	0.3465	0.3197	0.3744	0.3173	0.3736	0.4192
MAE:	0.5402	0.4359	0.5504	0.5360	0.5111	0.5693	0.5106	0.5642	0.6156
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6444	0.6681	0.6399	0.6452	0.6941	0.6601	0.6356	0.7150	0.4936
Std Y Pred:	0.0481	0.0529	0.0446	0.0536	0.0567	0.0504	0.0555	0.0604	0.0000
Avg Y Test:	0.6859	0.6084	0.6441	0.6198	0.7006	0.6722	0.6648	0.6122	0.4423
Std Y Test:	0.3629	0.3920	0.3850	0.3864	0.3293	0.3761	0.3675	0.3917	0.4434
R2 Score:	0.0164	-0.0117	0.0109	0.0254	0.0290	0.0222	0.0209	-0.0540	-0.0134
RMSE:	0.3600	0.3943	0.3828	0.3814	0.3245	0.3719	0.3636	0.4021	0.4463
MAE:	0.5566	0.5787	0.5760	0.5707	0.5145	0.5632	0.5594	0.5759	0.6473
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5866	0.6087	0.6081	0.5859	0.6304	0.6072	0.5758	0.6089	0.4236
Std Y Pred:	0.0441	0.0550	0.0451	0.0459	0.0585	0.0511	0.0434	0.0478	0.0000
Avg Y Test:	0.6056	0.5811	0.5921	0.5828	0.6308	0.6168	0.5855	0.6054	0.4081
Std Y Test:	0.3379	0.3386	0.3488	0.3510	0.3451	0.3338	0.3498	0.3487	0.4153
R2 Score:	0.0171	0.0245	0.0152	0.0315	0.0357	0.0225	0.0175	0.0064	-0.0014
RMSE:	0.3350	0.3344	0.3461	0.3454	0.3388	0.3300	0.3467	0.3476	0.4156
MAE:	0.5302	0.5267	0.5410	0.5411	0.5315	0.5292	0.5414	0.5436	0.6149

Table 41. NDCG@10 – Cluster 2

Cluster 3									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4518	0.6464	0.4473	0.4709	0.4631	0.3987	0.5093	0.4620	0.4897
Std Y Pred:	0.0495	0.0538	0.0454	0.0522	0.0560	0.0501	0.0569	0.0538	0.0027
Avg Y Test:	0.5119	0.6464	0.4562	0.4454	0.4750	0.4393	0.4827	0.4703	0.4810
Std Y Test:	0.4006	0.3271	0.4134	0.4054	0.4019	0.4127	0.4083	0.4123	0.4235
R2 Score:	-0.0129	0.0352	0.0117	0.0202	0.0242	0.0073	0.0236	0.0009	0.0005
RMSE:	0.4031	0.3213	0.4110	0.4013	0.3970	0.4111	0.4034	0.4121	0.4234
MAE:	0.5985	0.5123	0.6084	0.5982	0.5941	0.6058	0.5988	0.6079	0.6220
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4641	0.4869	0.4869	0.4605	0.4884	0.4844	0.4384	0.4940	0.4936
Std Y Pred:	0.0472	0.0517	0.0423	0.0511	0.0550	0.0468	0.0563	0.0646	0.0000
Avg Y Test:	0.4738	0.4217	0.4201	0.4585	0.4664	0.4797	0.4762	0.4619	0.4610
Std Y Test:	0.4145	0.4125	0.4159	0.4181	0.4123	0.4187	0.4226	0.4362	0.4362
R2 Score:	0.0153	-0.0148	-0.0072	0.0025	0.0261	0.0272	0.0076	0.0193	-0.0056
RMSE:	0.4113	0.4156	0.4173	0.4176	0.4069	0.4130	0.4210	0.4319	0.4374
MAE:	0.6092	0.6131	0.6161	0.6137	0.6047	0.6113	0.6176	0.6336	0.6378
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4294	0.4254	0.4436	0.4244	0.4440	0.4229	0.4251	0.4448	0.4236
Std Y Pred:	0.0449	0.0515	0.0436	0.0465	0.0532	0.0532	0.0425	0.0489	0.0000
Avg Y Test:	0.4354	0.4519	0.4263	0.4145	0.4445	0.4239	0.4019	0.4681	0.4275
Std Y Test:	0.3993	0.3862	0.3914	0.3920	0.3744	0.3958	0.4037	0.3905	0.4173
R2 Score:	0.0054	0.0244	0.0206	0.0149	0.0246	0.0248	0.0251	0.0244	-0.0001
RMSE:	0.3982	0.3815	0.3874	0.3890	0.3698	0.3909	0.3986	0.3857	0.4173
MAE:	0.5964	0.5751	0.5867	0.5862	0.5628	0.5884	0.5983	0.5815	0.6178

Table 42. NDCG@10 – Cluster 3

Cluster 4									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2861	0.4525	0.2878	0.2923	0.2810	0.2442	0.3188	0.2763	0.4802
Std Y Pred:	0.0477	0.0590	0.0489	0.0485	0.0521	0.0381	0.0513	0.0505	0.0030
Avg Y Test:	0.3374	0.4556	0.2662	0.2894	0.2781	0.2742	0.2914	0.2981	0.4002
Std Y Test:	0.4068	0.3906	0.3955	0.3933	0.3782	0.3937	0.3993	0.4024	0.4243
R2 Score:	-0.0135	0.0143	-0.0054	0.0089	0.0198	0.0068	0.0045	0.0075	-0.0345
RMSE:	0.4096	0.3878	0.3965	0.3916	0.3744	0.3924	0.3984	0.4009	0.4315
MAE:	0.5955	0.5840	0.5903	0.5855	0.5679	0.5755	0.5965	0.5905	0.6327
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3091	0.3135	0.3435	0.2864	0.3069	0.3237	0.2624	0.2637	0.4936
Std Y Pred:	0.0467	0.0486	0.0455	0.0502	0.0541	0.0481	0.0487	0.0591	0.0000
Avg Y Test:	0.3043	0.2815	0.2790	0.2873	0.2815	0.3045	0.2815	0.3515	0.4891
Std Y Test:	0.4097	0.4032	0.4032	0.4047	0.3980	0.4120	0.4036	0.4233	0.4363
R2 Score:	0.0166	-0.0120	-0.0119	-0.0080	0.0317	0.0061	0.0150	-0.0395	-0.0001
RMSE:	0.4063	0.4056	0.4056	0.4063	0.3917	0.4107	0.4005	0.4315	0.4364
MAE:	0.5996	0.6007	0.6077	0.5951	0.5884	0.6070	0.5844	0.6065	0.6367
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2785	0.2593	0.2975	0.2649	0.2584	0.2404	0.2785	0.2748	0.4236
Std Y Pred:	0.0442	0.0427	0.0437	0.0434	0.0523	0.0536	0.0452	0.0479	0.0000
Avg Y Test:	0.2658	0.2583	0.2846	0.2662	0.2512	0.2693	0.2497	0.3219	0.4438
Std Y Test:	0.3824	0.3798	0.3903	0.3848	0.3657	0.3914	0.3839	0.3954	0.4099
R2 Score:	0.0032	0.0206	0.0084	-0.0027	0.0200	0.0016	-0.0082	-0.0004	-0.0024
RMSE:	0.3818	0.3758	0.3887	0.3853	0.3620	0.3911	0.3854	0.3955	0.4104
MAE:	0.5742	0.5636	0.5824	0.5728	0.5532	0.5697	0.5792	0.5824	0.6078

Table 43. NDCG@10 – Cluster 4

Cluster 5									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.1200	0.2472	0.1032	0.1243	0.0977	0.1239	0.1497	0.1087	0.4680
Std Y Pred:	0.0436	0.0515	0.0619	0.0545	0.0612	0.0426	0.0638	0.0469	0.0056
Avg Y Test:	0.1652	0.2840	0.1407	0.1529	0.1419	0.1438	0.1513	0.1682	0.3186
Std Y Test:	0.3403	0.3938	0.3233	0.3354	0.3308	0.3317	0.3318	0.3446	0.4200
R2 Score:	-0.0085	0.0146	-0.0221	0.0061	-0.0111	0.0150	0.0299	-0.0127	-0.1233
RMSE:	0.3418	0.3909	0.3269	0.3344	0.3327	0.3292	0.3268	0.3467	0.4452
MAE:	0.4771	0.5721	0.4502	0.4721	0.4520	0.4686	0.4830	0.4718	0.6515
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.1435	0.1511	0.1690	0.1176	0.1192	0.1585	0.0967	0.0777	0.4936
Std Y Pred:	0.0467	0.0512	0.0578	0.0499	0.0662	0.0527	0.0590	0.0455	0.0000
Avg Y Test:	0.1917	0.1523	0.1437	0.1515	0.1506	0.1700	0.1400	0.2038	0.5251
Std Y Test:	0.3686	0.3422	0.3365	0.3392	0.3401	0.3558	0.3272	0.3737	0.4330
R2 Score:	0.0186	0.0085	0.0133	0.0019	0.0136	0.0210	-0.0004	-0.0934	-0.0053
RMSE:	0.3651	0.3407	0.3342	0.3389	0.3378	0.3520	0.3272	0.3907	0.4341
MAE:	0.5087	0.4946	0.5002	0.4702	0.4717	0.5072	0.4445	0.4844	0.6349
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.1319	0.1232	0.1449	0.1215	0.0829	0.0557	0.1105	0.1132	0.4236
Std Y Pred:	0.0414	0.0508	0.0455	0.0490	0.0613	0.0507	0.0612	0.0473	0.0000
Avg Y Test:	0.1616	0.1533	0.1535	0.1483	0.1462	0.1319	0.1207	0.1827	0.3616
Std Y Test:	0.3376	0.3357	0.3315	0.3264	0.3243	0.3117	0.2969	0.3516	0.4285
R2 Score:	0.0001	0.0035	0.0118	0.0116	-0.0120	-0.0495	0.0058	-0.0230	-0.0209
RMSE:	0.3376	0.3351	0.3296	0.3245	0.3263	0.3193	0.2960	0.3556	0.4329
MAE:	0.4828	0.4726	0.4838	0.4652	0.4402	0.4034	0.4355	0.4828	0.6370

Table 44. NDCG@10 – Cluster 5

6.3.1 RR@5 Clusters

Cluster 1									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6418	0.5797	0.6379	0.6014	0.6639	0.6059	0.6141	0.7157	0.6611
Std Y Pred:	0.0224	0.0197	0.0233	0.0189	0.0352	0.0162	0.0195	0.0435	0.0310
Avg Y Test:	0.5968	0.5639	0.6021	0.5951	0.6230	0.5993	0.6078	0.5791	0.6129
Std Y Test:	0.3055	0.3239	0.3262	0.3208	0.3351	0.3101	0.3276	0.3011	0.3112
R2 Score:	-0.0252	-0.0043	-0.0089	0.0058	-0.0144	0.0052	0.0052	-0.2095	-0.0288
RMSE:	0.3093	0.3246	0.3277	0.3199	0.3376	0.3093	0.3268	0.3311	0.3156
MAE:	0.5208	0.5157	0.5362	0.5204	0.5474	0.5119	0.5307	0.5466	0.5309
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6204	0.6091	0.6365	0.6391	0.6557	0.6232	0.6232	0.7655	0.6165
Std Y Pred:	0.0243	0.0188	0.0263	0.0218	0.0322	0.0220	0.0203	0.0481	0.0197
Avg Y Test:	0.5883	0.5986	0.6030	0.6111	0.6212	0.6038	0.6186	0.5842	0.6040
Std Y Test:	0.3093	0.3213	0.3298	0.3209	0.3279	0.3149	0.3408	0.3147	0.3206
R2 Score:	0.0011	0.0006	-0.0064	-0.0069	-0.0064	0.0034	0.0023	-0.3402	0.0033
RMSE:	0.3091	0.3212	0.3308	0.3220	0.3289	0.3143	0.3404	0.3643	0.3201
MAE:	0.5149	0.5233	0.5386	0.5326	0.5389	0.5219	0.5461	0.5703	0.5251
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6418	0.6239	0.6503	0.6213	0.6615	0.6459	0.6291	0.7346	0.6182
Std Y Pred:	0.0230	0.0216	0.0199	0.0195	0.0358	0.0230	0.0202	0.0412	0.0162
Avg Y Test:	0.5912	0.6115	0.5919	0.5996	0.6189	0.6184	0.6156	0.5716	0.5709
Std Y Test:	0.3091	0.3163	0.3250	0.3139	0.3367	0.3369	0.3117	0.3140	0.3354
R2 Score:	-0.0235	0.0027	-0.0321	-0.0004	-0.0043	-0.0031	0.0012	-0.2810	-0.0204
RMSE:	0.3127	0.3158	0.3302	0.3139	0.3374	0.3375	0.3115	0.3554	0.3387
MAE:	0.5225	0.5236	0.5391	0.5202	0.5471	0.5476	0.5217	0.5645	0.5405

Table 45. RR@5 - Cluster 1

Cluster 2									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5822	0.5337	0.5748	0.5511	0.5815	0.5586	0.5577	0.6061	0.5873
Std Y Pred:	0.0170	0.0120	0.0184	0.0150	0.0216	0.0143	0.0170	0.0286	0.0192
Avg Y Test:	0.5700	0.5630	0.5614	0.5516	0.5827	0.5556	0.5553	0.5558	0.5456
Std Y Test:	0.2969	0.3114	0.2908	0.3039	0.2953	0.3043	0.2890	0.2793	0.3012
R2 Score:	-0.0008	-0.0121	0.0013	0.0008	-0.0025	0.0004	0.0020	-0.0353	-0.0289
RMSE:	0.2970	0.3133	0.2906	0.3038	0.2956	0.3042	0.2887	0.2842	0.3056
MAE:	0.4928	0.4872	0.4840	0.4847	0.4890	0.4907	0.4727	0.4903	0.5033
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5570	0.5522	0.5721	0.5816	0.5812	0.5645	0.5662	0.6395	0.5658
Std Y Pred:	0.0174	0.0168	0.0183	0.0169	0.0195	0.0173	0.0169	0.0341	0.0148
Avg Y Test:	0.5585	0.5506	0.5756	0.5560	0.5852	0.5514	0.5612	0.5418	0.5692
Std Y Test:	0.2969	0.2966	0.3055	0.3028	0.3005	0.2972	0.2896	0.2825	0.3016
R2 Score:	-0.0021	0.0040	0.0016	-0.0035	0.0052	0.0006	-0.0025	-0.1370	-0.0066
RMSE:	0.2972	0.2960	0.3052	0.3033	0.2997	0.2972	0.2900	0.3012	0.3026
MAE:	0.4818	0.4788	0.4967	0.4977	0.4944	0.4863	0.4761	0.5151	0.4910
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5798	0.5641	0.5912	0.5671	0.5767	0.5839	0.5753	0.6274	0.5754
Std Y Pred:	0.0175	0.0173	0.0184	0.0157	0.0221	0.0186	0.0165	0.0280	0.0128
Avg Y Test:	0.5600	0.5704	0.5572	0.5522	0.5711	0.5635	0.5640	0.5668	0.5590
Std Y Test:	0.3039	0.2888	0.3001	0.2996	0.2901	0.2923	0.2985	0.2910	0.3045
R2 Score:	-0.0019	-0.0001	-0.0101	0.0011	0.0001	-0.0040	-0.0020	-0.0403	-0.0063
RMSE:	0.3042	0.2888	0.3016	0.2994	0.2900	0.2929	0.2988	0.2968	0.3055
MAE:	0.4986	0.4763	0.5006	0.4885	0.4826	0.4879	0.4914	0.5092	0.4981

Table 46. RR@5 – Cluster 2

Cluster 3									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5234	0.4939	0.5124	0.4970	0.5133	0.5060	0.4948	0.5103	0.5242
Std Y Pred:	0.0171	0.0115	0.0173	0.0155	0.0198	0.0152	0.0173	0.0262	0.0179
Avg Y Test:	0.5080	0.5025	0.5050	0.5041	0.5111	0.4915	0.5083	0.5008	0.5104
Std Y Test:	0.2621	0.2492	0.2557	0.2514	0.2539	0.2489	0.2510	0.2570	0.2608
R2 Score:	-0.0021	0.0021	0.0038	0.0016	0.0144	-0.0030	-0.0053	-0.0075	-0.0012
RMSE:	0.2624	0.2489	0.2552	0.2512	0.2520	0.2493	0.2516	0.2580	0.2609
MAE:	0.4296	0.4094	0.4190	0.4114	0.4156	0.4097	0.4076	0.4259	0.4300
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4973	0.4926	0.5097	0.5213	0.5185	0.5036	0.5059	0.5333	0.5123
Std Y Pred:	0.0178	0.0169	0.0183	0.0171	0.0185	0.0167	0.0171	0.0293	0.0158
Avg Y Test:	0.4928	0.5060	0.5254	0.5130	0.5169	0.4835	0.5063	0.4931	0.5234
Std Y Test:	0.2552	0.2548	0.2662	0.2571	0.2596	0.2366	0.2459	0.2533	0.2757
R2 Score:	0.0074	0.0045	-0.0024	0.0055	0.0077	0.0050	-0.0036	-0.0296	0.0008
RMSE:	0.2542	0.2543	0.2666	0.2564	0.2586	0.2360	0.2463	0.2571	0.2756
MAE:	0.4150	0.4135	0.4275	0.4244	0.4220	0.3989	0.4053	0.4386	0.4374
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5185	0.5031	0.5207	0.5099	0.5056	0.5191	0.5149	0.5363	0.5274
Std Y Pred:	0.0175	0.0176	0.0197	0.0158	0.0200	0.0177	0.0173	0.0251	0.0134
Avg Y Test:	0.5096	0.5123	0.5111	0.5036	0.5106	0.5018	0.5068	0.5021	0.5181
Std Y Test:	0.2579	0.2702	0.2568	0.2516	0.2540	0.2479	0.2572	0.2598	0.2442
R2 Score:	0.0026	0.0093	-0.0020	0.0049	0.0064	-0.0011	0.0070	-0.0140	0.0055
RMSE:	0.2576	0.2689	0.2570	0.2510	0.2531	0.2480	0.2563	0.2616	0.2435
MAE:	0.4231	0.4283	0.4267	0.4135	0.4143	0.4129	0.4195	0.4429	0.4161

Table 47. RR@5 - Cluster 3

Cluster 4									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4654	0.4559	0.4539	0.4436	0.4464	0.4555	0.4345	0.4231	0.4619
Std Y Pred:	0.0160	0.0111	0.0158	0.0145	0.0182	0.0137	0.0160	0.0260	0.0179
Avg Y Test:	0.4480	0.4422	0.4439	0.4470	0.4570	0.4409	0.4404	0.4600	0.4563
Std Y Test:	0.2116	0.1902	0.1958	0.1913	0.2168	0.1888	0.1898	0.2376	0.2188
R2 Score:	-0.0015	-0.0004	0.0037	0.0062	0.0033	-0.0034	0.0032	-0.0337	0.0040
RMSE:	0.2117	0.1902	0.1955	0.1907	0.2164	0.1891	0.1895	0.2416	0.2184
MAE:	0.3853	0.3674	0.3699	0.3653	0.3870	0.3655	0.3660	0.4111	0.3891
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4364	0.4371	0.4450	0.4623	0.4542	0.4493	0.4472	0.4336	0.4585
Std Y Pred:	0.0171	0.0151	0.0184	0.0162	0.0173	0.0159	0.0159	0.0305	0.0136
Avg Y Test:	0.4532	0.4447	0.4512	0.4496	0.4600	0.4292	0.4503	0.4440	0.4647
Std Y Test:	0.2073	0.1858	0.1990	0.2013	0.2130	0.1862	0.1962	0.2109	0.2186
R2 Score:	-0.0017	0.0047	0.0045	0.0011	0.0068	-0.0110	0.0096	0.0022	0.0061
RMSE:	0.2075	0.1853	0.1986	0.2012	0.2123	0.1872	0.1953	0.2107	0.2180
MAE:	0.3812	0.3613	0.3717	0.3717	0.3821	0.3687	0.3681	0.3842	0.3859
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4598	0.4410	0.4541	0.4569	0.4377	0.4602	0.4542	0.4503	0.4841
Std Y Pred:	0.0172	0.0166	0.0185	0.0150	0.0195	0.0162	0.0165	0.0251	0.0119
Avg Y Test:	0.4482	0.4558	0.4476	0.4381	0.4427	0.4569	0.4467	0.4535	0.4521
Std Y Test:	0.2092	0.2143	0.1941	0.1920	0.1995	0.2028	0.2001	0.2223	0.2020
R2 Score:	0.0062	-0.0009	-0.0018	-0.0009	0.0130	0.0134	0.0072	-0.0119	-0.0261
RMSE:	0.2085	0.2144	0.1942	0.1921	0.1982	0.2015	0.1994	0.2236	0.2046
MAE:	0.3823	0.3856	0.3667	0.3688	0.3725	0.3720	0.3732	0.3958	0.3737

Table 48. RR@5 - Cluster 4

Cluster 5									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4105	0.4157	0.4021	0.3936	0.3865	0.4092	0.3816	0.3272	0.3967
Std Y Pred:	0.0201	0.0159	0.0188	0.0183	0.0243	0.0163	0.0187	0.0377	0.0271
Avg Y Test:	0.4111	0.4181	0.3952	0.3958	0.3922	0.3884	0.4020	0.4332	0.4129
Std Y Test:	0.1899	0.1660	0.1618	0.1486	0.1779	0.1450	0.1765	0.2222	0.1774
R2 Score:	0.0025	-0.0018	-0.0076	0.0047	0.0031	-0.0233	-0.0091	-0.2513	-0.0213
RMSE:	0.1896	0.1661	0.1624	0.1483	0.1776	0.1467	0.1773	0.2486	0.1793
MAE:	0.3697	0.3504	0.3477	0.3335	0.3562	0.3391	0.3506	0.4055	0.3570
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3798	0.3871	0.3838	0.4060	0.3957	0.3897	0.3932	0.3201	0.4177
Std Y Pred:	0.0214	0.0175	0.0214	0.0199	0.0232	0.0221	0.0179	0.0462	0.0111
Avg Y Test:	0.4031	0.3891	0.4031	0.3981	0.3999	0.3887	0.3966	0.4464	0.4132
Std Y Test:	0.1718	0.1496	0.1725	0.1601	0.1874	0.1546	0.1751	0.2197	0.1735
R2 Score:	-0.0156	0.0168	-0.0120	0.0015	0.0020	0.0042	0.0000	-0.3520	0.0139
RMSE:	0.1732	0.1483	0.1735	0.1600	0.1872	0.1543	0.1751	0.2555	0.1723
MAE:	0.3475	0.3327	0.3489	0.3458	0.3657	0.3357	0.3545	0.4144	0.3566
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4016	0.3864	0.3936	0.4052	0.3681	0.4063	0.3972	0.3547	0.4451
Std Y Pred:	0.0232	0.0189	0.0207	0.0207	0.0278	0.0206	0.0201	0.0373	0.0144
Avg Y Test:	0.4078	0.3982	0.3947	0.3969	0.3874	0.3930	0.3951	0.4228	0.4143
Std Y Test:	0.1712	0.1748	0.1445	0.1626	0.1828	0.1733	0.1810	0.1928	0.1664
R2 Score:	-0.0104	0.0161	0.0037	-0.0095	-0.0138	-0.0124	-0.0055	-0.1575	-0.0234
RMSE:	0.1721	0.1734	0.1442	0.1634	0.1840	0.1744	0.1815	0.2074	0.1684
MAE:	0.3540	0.3506	0.3298	0.3493	0.3554	0.3587	0.3620	0.3725	0.3565

Table 49. RR@5 – Cluster 5

6.3.2 RR@10 Clusters

Cluster 1									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5971	0.5577	0.5845	0.5916	0.6328	0.5856	0.5888	0.7058	0.5906
Std Y Pred:	0.0260	0.0199	0.0207	0.0207	0.0349	0.0257	0.0275	0.0481	0.0178
Avg Y Test:	0.5671	0.5681	0.5876	0.5788	0.6224	0.5756	0.5861	0.5885	0.5785
Std Y Test:	0.3318	0.3256	0.3238	0.3203	0.3289	0.3234	0.3306	0.3073	0.3281
R2 Score:	-0.0075	-0.0024	-0.0077	-0.0035	0.0032	-0.0064	-0.0070	-0.1529	-0.0033
RMSE:	0.3330	0.3260	0.3250	0.3209	0.3284	0.3245	0.3318	0.3300	0.3286
MAE:	0.5293	0.5088	0.5185	0.5161	0.5355	0.5174	0.5264	0.5461	0.5232
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6147	0.6452	0.5693	0.5986	0.6320	0.5974	0.6030	0.6620	0.4485
Std Y Pred:	0.0291	0.0362	0.0220	0.0249	0.0389	0.0217	0.0260	0.0511	0.0000
Avg Y Test:	0.5794	0.5739	0.5713	0.5720	0.6278	0.5777	0.5813	0.5759	0.4452
Std Y Test:	0.3131	0.3147	0.3062	0.2989	0.3230	0.3268	0.3284	0.3226	0.2208
R2 Score:	-0.0066	-0.0462	-0.0028	-0.0053	-0.0009	-0.0047	-0.0026	-0.1058	-0.0002
RMSE:	0.3142	0.3219	0.3066	0.2997	0.3231	0.3276	0.3288	0.3392	0.2208
MAE:	0.5164	0.5301	0.4949	0.4992	0.5315	0.5250	0.5263	0.5449	0.3986
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5889	0.5622	0.5741	0.5748	0.6395	0.5994	0.6137	0.6608	0.5450
Std Y Pred:	0.0240	0.0206	0.0223	0.0208	0.0365	0.0228	0.0263	0.0408	0.0170
Avg Y Test:	0.5803	0.5637	0.5710	0.5739	0.6172	0.5781	0.5884	0.5583	0.5650
Std Y Test:	0.3189	0.3265	0.3302	0.3206	0.3327	0.3184	0.3097	0.2989	0.3285
R2 Score:	-0.0072	0.0018	-0.0002	0.0011	-0.0131	-0.0053	-0.0120	-0.1166	-0.0060
RMSE:	0.3201	0.3262	0.3302	0.3204	0.3349	0.3192	0.3116	0.3159	0.3295
MAE:	0.5148	0.5093	0.5211	0.5096	0.5441	0.5181	0.5160	0.5275	0.5084

Table 50. RR@10 – Cluster 1

Cluster 2									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5280	0.5044	0.5249	0.5290	0.5421	0.5112	0.5131	0.5854	0.5219
Std Y Pred:	0.0191	0.0154	0.0180	0.0182	0.0244	0.0209	0.0212	0.0332	0.0206
Avg Y Test:	0.5257	0.5153	0.5235	0.5342	0.5420	0.5292	0.5338	0.5267	0.5299
Std Y Test:	0.2583	0.2368	0.2491	0.2528	0.2635	0.2457	0.2496	0.2567	0.2362
R2 Score:	0.0007	0.0014	0.0065	0.0041	0.0008	-0.0035	-0.0076	-0.0466	0.0043
RMSE:	0.2582	0.2367	0.2483	0.2523	0.2634	0.2462	0.2506	0.2626	0.2357
MAE:	0.4259	0.3891	0.4112	0.4185	0.4329	0.4045	0.4070	0.4605	0.3982
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5312	0.5458	0.5108	0.5300	0.5360	0.5333	0.5286	0.5286	N/A
Std Y Pred:	0.0237	0.0284	0.0170	0.0204	0.0252	0.0189	0.0206	0.0347	N/A
Avg Y Test:	0.5310	0.5502	0.5307	0.5236	0.5266	0.5346	0.5526	0.5345	N/A
Std Y Test:	0.2613	0.2726	0.2604	0.2489	0.2590	0.2504	0.2670	0.2681	N/A
R2 Score:	0.0013	0.0014	-0.0058	0.0114	0.0023	0.0080	-0.0026	-0.0069	N/A
RMSE:	0.2612	0.2724	0.2611	0.2475	0.2587	0.2494	0.2673	0.2691	N/A
MAE:	0.4270	0.4487	0.4174	0.4141	0.4257	0.4172	0.4341	0.4395	N/A
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5229	0.5034	0.5096	0.5081	0.5454	0.5197	0.5298	0.5543	0.4975
Std Y Pred:	0.0200	0.0170	0.0188	0.0194	0.0246	0.0228	0.0229	0.0278	0.0142
Avg Y Test:	0.5303	0.5199	0.5253	0.5210	0.5450	0.5203	0.5274	0.5350	0.5129
Std Y Test:	0.2574	0.2386	0.2382	0.2346	0.2698	0.2433	0.2665	0.2710	0.2320
R2 Score:	-0.0009	0.0026	0.0038	0.0037	-0.0005	0.0125	0.0061	-0.0154	-0.0042
RMSE:	0.2575	0.2383	0.2377	0.2341	0.2698	0.2418	0.2657	0.2730	0.2325
MAE:	0.4208	0.3924	0.3920	0.3902	0.4424	0.4022	0.4317	0.4578	0.3833

Table 51. RR@10 – Cluster 2 – N/A occurs when values for the Trans. + MLP converged to a singular value preventing more than one cluster from forming.

Cluster 3									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4601	0.4479	0.4617	0.4652	0.4651	0.4402	0.4401	0.4798	0.4511
Std Y Pred:	0.0193	0.0161	0.0180	0.0186	0.0211	0.0189	0.0202	0.0292	0.0188
Avg Y Test:	0.4642	0.4515	0.4554	0.4630	0.4706	0.4409	0.4511	0.4564	0.4616
Std Y Test:	0.2036	0.1960	0.1795	0.1866	0.1981	0.1737	0.1744	0.2093	0.1823
R2 Score:	0.0064	0.0091	0.0047	0.0077	0.0165	0.0048	-0.0062	-0.0150	-0.0133
RMSE:	0.2030	0.1951	0.1791	0.1859	0.1965	0.1733	0.1750	0.2109	0.1835
MAE:	0.3712	0.3707	0.3482	0.3503	0.3578	0.3495	0.3482	0.3805	0.3525
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4500	0.4477	0.4479	0.3983	0.4547	0.3992	0.4593	0.4193	N/A
Std Y Pred:	0.0217	0.0235	0.0178	0.0176	0.0214	0.0174	0.0189	0.0280	N/A
Avg Y Test:	0.4751	0.4656	0.4479	0.3945	0.4565	0.4031	0.4630	0.4559	N/A
Std Y Test:	0.1941	0.1954	0.1866	0.1430	0.1839	0.1474	0.1804	0.2048	N/A
R2 Score:	0.0045	0.0038	0.0117	0.0199	0.0218	0.0105	0.0076	-0.0353	N/A
RMSE:	0.1937	0.1950	0.1855	0.1415	0.1819	0.1466	0.1797	0.2084	N/A
MAE:	0.3590	0.3623	0.3586	0.3266	0.3491	0.3278	0.3445	0.3792	N/A
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4524	0.4383	0.4443	0.4422	0.4674	0.4439	0.4544	0.4626	0.4471
Std Y Pred:	0.0200	0.0181	0.0173	0.0178	0.0209	0.0195	0.0210	0.0253	0.0149
Avg Y Test:	0.4544	0.4577	0.4544	0.4453	0.4743	0.4575	0.4705	0.4634	0.4543
Std Y Test:	0.1822	0.1966	0.1807	0.1756	0.1948	0.1831	0.1944	0.2136	0.2018
R2 Score:	0.0081	0.0048	0.0082	0.0187	0.0129	0.0123	0.0124	0.0106	0.0086
RMSE:	0.1815	0.1961	0.1799	0.1740	0.1936	0.1819	0.1932	0.2125	0.2010
MAE:	0.3513	0.3698	0.3530	0.3492	0.3513	0.3542	0.3574	0.3793	0.3754

Table 52. RR@10 – Cluster 3 – N/A occurs when values for the Trans. + MLP converged to a singular value preventing more than one cluster from forming.

Cluster 4									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3942	0.3962	0.4039	0.4030	0.3967	0.3793	0.3745	0.3845	0.3907
Std Y Pred:	0.0176	0.0128	0.0162	0.0164	0.0193	0.0169	0.0177	0.0263	0.0163
Avg Y Test:	0.3902	0.3922	0.3953	0.4021	0.3998	0.3853	0.3904	0.3949	0.4103
Std Y Test:	0.1412	0.1503	0.1410	0.1444	0.1473	0.1396	0.1378	0.1574	0.1520
R2 Score:	0.0156	0.0134	0.0101	0.0186	0.0149	0.0216	0.0129	-0.0150	-0.0032
RMSE:	0.1401	0.1493	0.1403	0.1431	0.1462	0.1381	0.1369	0.1585	0.1523
MAE:	0.3275	0.3373	0.3272	0.3271	0.3276	0.3171	0.3145	0.3370	0.3303
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3813	0.3772	0.3886	0.4613	0.3857	0.4631	0.3972	0.3283	N/A
Std Y Pred:	0.0185	0.0198	0.0169	0.0192	0.0206	0.0196	0.0171	0.0257	N/A
Avg Y Test:	0.3959	0.3957	0.3894	0.4509	0.3870	0.4599	0.4039	0.3973	N/A
Std Y Test:	0.1419	0.1375	0.1409	0.1940	0.1408	0.1868	0.1486	0.1565	N/A
R2 Score:	0.0004	-0.0030	0.0139	0.0056	0.0261	0.0178	0.0190	-0.1919	N/A
RMSE:	0.1419	0.1377	0.1400	0.1935	0.1389	0.1851	0.1472	0.1709	N/A
MAE:	0.3181	0.3140	0.3221	0.3632	0.3206	0.3514	0.3292	0.3335	N/A
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3839	0.3794	0.3864	0.3867	0.3988	0.3849	0.3828	0.3799	0.3944
Std Y Pred:	0.0188	0.0145	0.0154	0.0151	0.0200	0.0156	0.0195	0.0234	0.0131
Avg Y Test:	0.3965	0.3950	0.3922	0.3905	0.4044	0.3982	0.4069	0.3980	0.3962
Std Y Test:	0.1428	0.1423	0.1411	0.1374	0.1543	0.1520	0.1520	0.1603	0.1486
R2 Score:	0.0133	0.0094	0.0106	0.0215	0.0047	0.0108	-0.0112	-0.0088	0.0085
RMSE:	0.1419	0.1416	0.1403	0.1359	0.1540	0.1511	0.1529	0.1610	0.1480
MAE:	0.3195	0.3217	0.3209	0.3187	0.3361	0.3273	0.3264	0.3340	0.3355

Table 53. RR@10 – Cluster 4 – N/A occurs when values for the Trans. + MLP converged to a singular value preventing more than one cluster from forming.

Cluster 5									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3360	0.3551	0.3456	0.3470	0.3301	0.3213	0.3150	0.2928	0.3390
Std Y Pred:	0.0207	0.0138	0.0206	0.0183	0.0210	0.0225	0.0233	0.0338	0.0145
Avg Y Test:	0.3477	0.3543	0.3327	0.3344	0.3263	0.3354	0.3321	0.3543	0.3465
Std Y Test:	0.1189	0.1135	0.1072	0.1110	0.1090	0.1062	0.1059	0.1331	0.1175
R2 Score:	0.0072	0.0070	-0.0314	-0.0043	0.0286	-0.0316	-0.0419	-0.2367	0.0186
RMSE:	0.1184	0.1131	0.1088	0.1112	0.1075	0.1079	0.1081	0.1480	0.1164
MAE:	0.2815	0.2847	0.2859	0.2868	0.2721	0.2737	0.2675	0.3035	0.2811
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3217	0.3080	0.3285	0.3355	0.3128	0.3388	0.3388	0.2370	N/A
Std Y Pred:	0.0203	0.0237	0.0208	0.0206	0.0252	0.0194	0.0196	0.0330	N/A
Avg Y Test:	0.3391	0.3318	0.3380	0.3306	0.3190	0.3348	0.3350	0.3592	N/A
Std Y Test:	0.1218	0.1102	0.1165	0.1114	0.1030	0.1105	0.1076	0.1278	N/A
R2 Score:	-0.0141	-0.0346	-0.0002	-0.0095	0.0303	0.0097	0.0100	-0.9706	N/A
RMSE:	0.1226	0.1121	0.1165	0.1119	0.1015	0.1100	0.1070	0.1794	N/A
MAE:	0.2819	0.2689	0.2804	0.2779	0.2544	0.2810	0.2777	0.3548	N/A
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.3178	0.3342	0.3328	0.3337	0.3337	0.3252	0.3323	0.3193	0.3537
Std Y Pred:	0.0218	0.0159	0.0179	0.0185	0.0185	0.0261	0.0168	0.0192	0.0124
Avg Y Test:	0.3380	0.3490	0.3396	0.3322	0.3322	0.3242	0.3376	0.3376	0.3623
Std Y Test:	0.1158	0.1203	0.1155	0.1084	0.1084	0.1005	0.1073	0.1099	0.1292
R2 Score:	-0.0096	0.0075	0.0202	0.0000	0.0000	0.0282	0.0085	-0.0137	0.0025
RMSE:	0.1164	0.1198	0.1143	0.1084	0.1084	0.0991	0.1069	0.1106	0.1290
MAE:	0.2756	0.2794	0.2786	0.2762	0.2762	0.2598	0.2742	0.2721	0.2936

Table 54. RR@10 – Cluster 5 – N/A occurs when values for the Trans. + MLP converged to a singular value preventing more than one cluster from forming.

6.4.1 ERR@5 Clusters.

Cluster 1									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.8539	0.6573	0.8574	0.8713	0.9552	0.8491	0.8685	0.9050	0.8453
Std Y Pred:	0.0347	0.0001	0.0326	0.0324	0.0649	0.0313	0.0464	0.0370	0.0279
Avg Y Test:	0.8500	0.7857	0.8542	0.8558	0.9039	0.8545	0.8664	0.8426	0.8513
Std Y Test:	0.4645	0.4923	0.4612	0.4668	0.4992	0.4659	0.4698	0.4708	0.4777
R2 Score:	0.0070	-0.0680	0.0075	0.0089	0.0006	0.0122	0.0078	-0.0141	0.0096
RMSE:	0.4628	0.5087	0.4594	0.4647	0.4991	0.4630	0.4679	0.4741	0.4754
MAE:	0.5473	0.5721	0.5428	0.5503	0.5923	0.5452	0.5529	0.5662	0.5518
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.8121	0.8883	0.8497	0.8607	0.9218	0.8375	0.8645	0.8690	0.6508
Std Y Pred:	0.0317	0.0348	0.0329	0.0322	0.0577	0.0348	0.0430	0.0381	0.0001
Avg Y Test:	0.8558	0.8750	0.8609	0.8639	0.9127	0.8598	0.8720	0.8518	0.7814
Std Y Test:	0.4614	0.4706	0.4574	0.4679	0.5097	0.4701	0.4774	0.4733	0.4316
R2 Score:	-0.0064	0.0124	0.0077	0.0090	0.0093	0.0080	0.0066	0.0076	-0.0915
RMSE:	0.4629	0.4676	0.4556	0.4658	0.5073	0.4682	0.4758	0.4715	0.4509
MAE:	0.5418	0.5545	0.5373	0.5484	0.5806	0.5474	0.5536	0.5550	0.5492
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.8575	0.8526	0.8438	0.8436	0.9412	0.8586	0.8733	0.8837	0.6564
Std Y Pred:	0.0299	0.0390	0.0334	0.0318	0.0620	0.0315	0.0370	0.0364	0.0000
Avg Y Test:	0.8421	0.8813	0.8622	0.8615	0.9010	0.8582	0.8630	0.8482	0.6665
Std Y Test:	0.4547	0.4777	0.4688	0.4701	0.5037	0.4767	0.4808	0.4713	0.4645
R2 Score:	0.0060	0.0080	0.0061	0.0024	0.0028	0.0091	0.0046	0.0009	-0.0005
RMSE:	0.4533	0.4758	0.4674	0.4695	0.5030	0.4745	0.4797	0.4711	0.4646
MAE:	0.5434	0.5521	0.5464	0.5483	0.5890	0.5520	0.5594	0.5599	0.5750

Table 55. ERR@5 – Cluster 1

Cluster 2									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7311	0.6575	0.7370	0.7489	0.8019	0.7309	0.7250	0.7721	0.7268
Std Y Pred:	0.0385	0.0001	0.0381	0.0403	0.0429	0.0378	0.0442	0.0434	0.0363
Avg Y Test:	0.7543	0.8464	0.7247	0.7348	0.8195	0.7295	0.7331	0.7218	0.7288
Std Y Test:	0.4634	0.4423	0.4323	0.4366	0.4573	0.4368	0.4387	0.4343	0.4127
R2 Score:	-0.0014	-0.1825	0.0073	0.0139	0.0075	0.0129	0.0112	-0.0016	0.0166
RMSE:	0.4637	0.4810	0.4307	0.4336	0.4556	0.4340	0.4363	0.4347	0.4093
MAE:	0.5550	0.5605	0.5429	0.5395	0.5392	0.5425	0.5399	0.5481	0.5304
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7029	0.7639	0.7307	0.7448	0.7767	0.7135	0.7311	0.7383	0.6505
Std Y Pred:	0.0348	0.0400	0.0385	0.0377	0.0414	0.0387	0.0415	0.0408	0.0001
Avg Y Test:	0.7405	0.7693	0.7190	0.7372	0.7993	0.7339	0.7357	0.7423	0.7915
Std Y Test:	0.4469	0.4393	0.4266	0.4382	0.4373	0.4366	0.4359	0.4372	0.4804
R2 Score:	-0.0054	0.0001	0.0094	0.0113	-0.0006	0.0107	0.0057	0.0041	-0.0861
RMSE:	0.4481	0.4393	0.4246	0.4357	0.4374	0.4343	0.4346	0.4363	0.5007
MAE:	0.5492	0.5360	0.5384	0.5372	0.5334	0.5384	0.5402	0.5406	0.5698
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7477	0.7292	0.7222	0.7371	0.7906	0.7356	0.7412	0.7541	N/A
Std Y Pred:	0.0376	0.0376	0.0389	0.0346	0.0426	0.0389	0.0415	0.0419	N/A
Avg Y Test:	0.7397	0.7545	0.7261	0.7417	0.8176	0.7268	0.7248	0.7196	N/A
Std Y Test:	0.4558	0.4316	0.4349	0.4412	0.4528	0.4143	0.4100	0.4410	N/A
R2 Score:	0.0047	0.0028	0.0018	0.0066	0.0025	0.0193	0.0142	0.0036	N/A
RMSE:	0.4547	0.4310	0.4345	0.4398	0.4522	0.4103	0.4071	0.4402	N/A
MAE:	0.5511	0.5315	0.5406	0.5405	0.5370	0.5319	0.5321	0.5459	N/A

Table 56. ERR@5 – Cluster 2 – N/A occurs when values for the Trans. + MLP converged to a singular value preventing more than one cluster from forming.

Cluster 3									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5914	0.6570	0.5887	0.5935	0.6492	0.5874	0.5624	0.6004	0.5988
Std Y Pred:	0.0414	0.0001	0.0445	0.0450	0.0477	0.0407	0.0468	0.0546	0.0355
Avg Y Test:	0.6012	0.6339	0.5897	0.6179	0.6699	0.6094	0.6032	0.6122	0.6179
Std Y Test:	0.4061	0.4051	0.4064	0.4237	0.4155	0.4163	0.4053	0.4001	0.4183
R2 Score:	0.0097	-0.0032	0.0152	0.0144	0.0167	0.0088	0.0088	0.0099	0.0059
RMSE:	0.4041	0.4058	0.4033	0.4206	0.4120	0.4144	0.4035	0.3981	0.4171
MAE:	0.5356	0.5443	0.5317	0.5345	0.5339	0.5338	0.5268	0.5353	0.5351
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5690	0.6138	0.5794	0.5974	0.6305	0.5682	0.5762	0.5803	0.6502
Std Y Pred:	0.0406	0.0455	0.0461	0.0426	0.0444	0.0433	0.0470	0.0467	0.0001
Avg Y Test:	0.6118	0.6316	0.6269	0.6112	0.6348	0.6068	0.5905	0.6030	0.6699
Std Y Test:	0.4240	0.4176	0.4473	0.4189	0.4030	0.4094	0.3959	0.4025	0.4586
R2 Score:	0.0016	0.0007	0.0018	0.0130	0.0149	0.0006	0.0120	0.0097	-0.0018
RMSE:	0.4237	0.4174	0.4469	0.4161	0.4000	0.4093	0.3935	0.4005	0.4590
MAE:	0.5355	0.5383	0.5484	0.5394	0.5300	0.5310	0.5289	0.5366	0.5690
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5967	0.5973	0.5764	0.5981	0.6402	0.5845	0.5776	0.5923	N/A
Std Y Pred:	0.0430	0.0418	0.0435	0.0418	0.0467	0.0434	0.0493	0.0506	N/A
Avg Y Test:	0.5775	0.6030	0.6106	0.6180	0.6562	0.6260	0.5897	0.5941	N/A
Std Y Test:	0.4134	0.3960	0.4120	0.4154	0.4076	0.4260	0.4114	0.4009	N/A
R2 Score:	0.0068	0.0110	0.0041	0.0158	0.0186	0.0092	0.0186	0.0043	N/A
RMSE:	0.4120	0.3938	0.4112	0.4122	0.4038	0.4240	0.4075	0.4000	N/A
MAE:	0.5404	0.5282	0.5301	0.5337	0.5312	0.5362	0.5303	0.5325	N/A

Table 57. ERR@5 – Cluster 3 – N/A occurs when values for the Trans. + MLP converged to a singular value preventing more than one cluster from forming.

Cluster 4									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4363	0.6568	0.4165	0.4303	0.4703	0.4352	0.3923	0.3898	0.4676
Std Y Pred:	0.0470	0.0001	0.0508	0.0493	0.0581	0.0463	0.0534	0.0623	0.0392
Avg Y Test:	0.4690	0.4924	0.4629	0.4757	0.4782	0.4536	0.4310	0.4772	0.4792
Std Y Test:	0.3619	0.4085	0.3879	0.3683	0.3597	0.3594	0.3577	0.4235	0.3400
R2 Score:	0.0179	-0.1620	0.0127	0.0026	0.0374	0.0261	0.0144	-0.0199	0.0088
RMSE:	0.3587	0.4404	0.3855	0.3678	0.3529	0.3547	0.3552	0.4277	0.3385
MAE:	0.5066	0.5889	0.5157	0.5121	0.5047	0.5023	0.5010	0.5340	0.5064
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4168	0.4498	0.4145	0.4417	0.4708	0.4057	0.4074	0.4055	0.6499
Std Y Pred:	0.0476	0.0526	0.0500	0.0491	0.0518	0.0499	0.0524	0.0548	0.0001
Avg Y Test:	0.4469	0.4937	0.4627	0.4751	0.4782	0.4332	0.4029	0.4733	0.5430
Std Y Test:	0.3661	0.3757	0.3774	0.3673	0.3688	0.3438	0.3397	0.3984	0.4238
R2 Score:	0.0123	0.0232	0.0033	0.0150	0.0329	0.0117	0.0202	-0.0124	-0.0636
RMSE:	0.3639	0.3714	0.3768	0.3645	0.3627	0.3418	0.3363	0.4009	0.4371
MAE:	0.5073	0.5101	0.5113	0.5086	0.5125	0.4934	0.4950	0.5215	0.5794
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4436	0.4384	0.4199	0.4399	0.4672	0.4241	0.3974	0.4057	N/A
Std Y Pred:	0.0447	0.0497	0.0481	0.0479	0.0552	0.0509	0.0535	0.0562	N/A
Avg Y Test:	0.4391	0.4524	0.4440	0.4497	0.4757	0.4618	0.4288	0.4770	N/A
Std Y Test:	0.3574	0.3817	0.3623	0.3536	0.3603	0.3599	0.3538	0.3588	N/A
R2 Score:	0.0062	0.0339	0.0238	0.0194	0.0273	0.0060	0.0278	-0.0373	N/A
RMSE:	0.3563	0.3752	0.3580	0.3501	0.3554	0.3588	0.3488	0.3654	N/A
MAE:	0.5121	0.5132	0.5069	0.5067	0.5081	0.5037	0.4964	0.5121	N/A

Table 58. ERR@5 - Cluster 4 – N/A occurs when values for the Trans. + MLP converged to a singular value preventing more than one cluster from forming.

Cluster 5									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2613	0.6565	0.2364	0.2401	0.2529	0.2607	0.1886	0.1587	0.3228
Std Y Pred:	0.0613	0.0001	0.0618	0.0694	0.0771	0.0651	0.0812	0.0686	0.0392
Avg Y Test:	0.2536	0.3510	0.2319	0.2239	0.2274	0.2344	0.2120	0.2883	0.2760
Std Y Test:	0.2666	0.3316	0.2950	0.2582	0.2609	0.2800	0.2644	0.3175	0.3260
R2 Score:	0.0960	-0.8486	0.0606	0.0832	0.0811	0.0673	0.0570	-0.1205	0.0374
RMSE:	0.2535	0.4509	0.2859	0.2472	0.2501	0.2704	0.2567	0.3361	0.3198
MAE:	0.4418	0.6322	0.4371	0.4318	0.4387	0.4487	0.4212	0.4783	0.4806
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2424	0.2477	0.2395	0.2489	0.2634	0.2262	0.2013	0.1984	0.6496
Std Y Pred:	0.0603	0.0717	0.0598	0.0743	0.0733	0.0625	0.0791	0.0680	0.0001
Avg Y Test:	0.2536	0.2604	0.2361	0.2277	0.2367	0.2128	0.1669	0.2809	0.3755
Std Y Test:	0.2846	0.2925	0.2695	0.2590	0.2622	0.2719	0.2268	0.2994	0.3564
R2 Score:	0.0724	0.0888	0.0659	0.0766	0.0887	0.0656	-0.0292	-0.0134	-0.5914
RMSE:	0.2741	0.2792	0.2605	0.2488	0.2503	0.2628	0.2301	0.3014	0.4496
MAE:	0.4492	0.4460	0.4403	0.4361	0.4386	0.4344	0.4153	0.4589	0.6280
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2760	0.2549	0.2435	0.2652	0.2511	0.2350	0.1988	0.1982	N/A
Std Y Pred:	0.0591	0.0651	0.0627	0.0650	0.0794	0.0676	0.0662	0.0661	N/A
Avg Y Test:	0.2304	0.2185	0.2253	0.2154	0.2165	0.2328	0.1994	0.2707	N/A
Std Y Test:	0.2742	0.2657	0.2683	0.2546	0.2488	0.2751	0.2557	0.3415	N/A
R2 Score:	0.0364	0.0534	0.0614	0.0505	0.0706	0.0827	0.0697	-0.0057	N/A
RMSE:	0.2692	0.2585	0.2599	0.2481	0.2399	0.2634	0.2467	0.3424	N/A
MAE:	0.4534	0.4414	0.4395	0.4391	0.4298	0.4373	0.4169	0.4705	N/A

Table 59. ERR@5 – Cluster 5 – N/A occurs when values for the Trans. + MLP converged to a singular value preventing more than one cluster from forming.

6.4.2 ERR@10 Clusters

Cluster 1									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.8568	0.8620	0.8240	0.8459	0.9591	0.8587	0.8516	0.8781	0.8072
Std Y Pred:	0.0340	0.0248	0.0319	0.0314	0.0686	0.0393	0.0314	0.0359	0.0273
Avg Y Test:	0.8568	0.8699	0.8577	0.8588	0.9162	0.8762	0.8574	0.8462	0.8400
Std Y Test:	0.4852	0.4748	0.4733	0.4713	0.4909	0.4858	0.4731	0.4646	0.4499
R2 Score:	0.0089	0.0039	0.0041	0.0113	0.0119	0.0049	0.0087	-0.0008	0.0013
RMSE:	0.4831	0.4739	0.4724	0.4687	0.4880	0.4846	0.4710	0.4648	0.4496
MAE:	0.5527	0.5427	0.5402	0.5389	0.5894	0.5472	0.5409	0.5535	0.5271
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.8403	0.8618	0.8242	0.8487	0.9557	0.8725	0.8696	0.8757	0.8130
Std Y Pred:	0.0282	0.0335	0.0319	0.0345	0.0688	0.0469	0.0410	0.0359	0.0245
Avg Y Test:	0.8604	0.8792	0.8695	0.8706	0.9058	0.8739	0.8820	0.8461	0.8526
Std Y Test:	0.4787	0.4756	0.4744	0.4789	0.4900	0.4777	0.4894	0.4579	0.4731
R2 Score:	-0.0030	0.0096	-0.0026	0.0064	0.0052	0.0025	0.0038	-0.0027	-0.0005
RMSE:	0.4794	0.4733	0.4750	0.4773	0.4887	0.4771	0.4885	0.4585	0.4732
MAE:	0.5454	0.5431	0.5388	0.5426	0.5835	0.5481	0.5497	0.5550	0.5384
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.8401	0.8411	0.8460	0.8328	0.9059	0.8824	0.8560	0.8800	0.7886
Std Y Pred:	0.0305	0.0432	0.0339	0.0316	0.0561	0.0411	0.0390	0.0373	0.0261
Avg Y Test:	0.8491	0.8761	0.8645	0.8651	0.8908	0.8789	0.8707	0.8314	0.8358
Std Y Test:	0.4483	0.4504	0.4700	0.4702	0.4733	0.4932	0.4788	0.4778	0.4603
R2 Score:	0.0052	0.0053	0.0052	0.0069	0.0230	0.0102	0.0108	-0.0070	-0.0023
RMSE:	0.4472	0.4492	0.4688	0.4685	0.4678	0.4907	0.4762	0.4795	0.4608
MAE:	0.5345	0.5314	0.5393	0.5367	0.5599	0.5568	0.5450	0.5586	0.5369

Table 60. ERR@10 – Cluster 1

Cluster 2									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7392	0.7398	0.7103	0.7367	0.7953	0.7385	0.7375	0.7521	0.7058
Std Y Pred:	0.0389	0.0351	0.0373	0.0375	0.0447	0.0380	0.0357	0.0412	0.0333
Avg Y Test:	0.7438	0.7274	0.7322	0.7341	0.8217	0.7512	0.7479	0.7252	0.7456
Std Y Test:	0.4179	0.4085	0.4189	0.4260	0.4517	0.4118	0.4270	0.4374	0.4673
R2 Score:	0.0114	-0.0032	0.0051	0.0159	-0.0085	0.0094	0.0134	0.0028	-0.0005
RMSE:	0.4155	0.4091	0.4178	0.4226	0.4536	0.4099	0.4241	0.4368	0.4674
MAE:	0.5297	0.5268	0.5299	0.5319	0.5266	0.5252	0.5329	0.5338	0.5525
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7423	0.7355	0.7114	0.7350	0.7998	0.7376	0.7475	0.7498	0.7267
Std Y Pred:	0.0330	0.0388	0.0344	0.0371	0.0414	0.0409	0.0372	0.0403	0.0276
Avg Y Test:	0.7288	0.7459	0.7256	0.7374	0.8232	0.7426	0.7406	0.7283	0.7292
Std Y Test:	0.4185	0.4079	0.4137	0.4161	0.4620	0.4209	0.3940	0.4449	0.4256
R2 Score:	0.0038	0.0066	0.0035	0.0108	-0.0051	0.0093	0.0136	0.0026	0.0083
RMSE:	0.4177	0.4066	0.4129	0.4139	0.4631	0.4190	0.3913	0.4443	0.4238
MAE:	0.5309	0.5158	0.5280	0.5272	0.5335	0.5292	0.5196	0.5371	0.5386
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.7351	0.7259	0.7298	0.7255	0.7625	0.7638	0.7376	0.7529	0.6938
Std Y Pred:	0.0340	0.0357	0.0366	0.0352	0.0406	0.0368	0.0379	0.0407	0.0324
Avg Y Test:	0.7514	0.7717	0.7250	0.7561	0.7908	0.7557	0.7407	0.7329	0.7314
Std Y Test:	0.4687	0.4468	0.4215	0.4383	0.4564	0.4066	0.4176	0.4388	0.4358
R2 Score:	0.0048	0.0007	0.0066	0.0016	0.0099	0.0124	0.0111	0.0167	-0.0011
RMSE:	0.4676	0.4467	0.4201	0.4379	0.4541	0.4041	0.4152	0.4351	0.4360
MAE:	0.5459	0.5321	0.5307	0.5354	0.5302	0.5237	0.5283	0.5438	0.5371

Table 61. ERR@10 – Cluster 2

Cluster 3									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.5881	0.6182	0.5711	0.5886	0.6484	0.5923	0.5989	0.5953	0.5768
Std Y Pred:	0.0458	0.0358	0.0402	0.0435	0.0448	0.0426	0.0415	0.0462	0.0342
Avg Y Test:	0.6188	0.6367	0.6019	0.6176	0.6616	0.6314	0.6189	0.6138	0.6190
Std Y Test:	0.3888	0.3823	0.3886	0.4014	0.3995	0.4174	0.3987	0.3991	0.3994
R2 Score:	-0.0068	0.0106	0.0054	0.0110	0.0174	0.0031	0.0080	0.0095	0.0048
RMSE:	0.3902	0.3803	0.3875	0.3992	0.3960	0.4168	0.3971	0.3972	0.3984
MAE:	0.5109	0.5139	0.5122	0.5236	0.5224	0.5303	0.5224	0.5274	0.5165
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6019	0.5916	0.5820	0.5908	0.6546	0.5825	0.6046	0.5948	0.6201
Std Y Pred:	0.0444	0.0423	0.0403	0.0432	0.0445	0.0468	0.0427	0.0465	0.0306
Avg Y Test:	0.5956	0.6357	0.6223	0.6129	0.6789	0.6080	0.6311	0.6012	0.6245
Std Y Test:	0.3975	0.4106	0.3881	0.3983	0.3989	0.3931	0.4200	0.4147	0.4015
R2 Score:	0.0181	-0.0083	-0.0079	0.0019	0.0088	0.0083	0.0116	-0.0018	0.0140
RMSE:	0.3939	0.4123	0.3896	0.3979	0.3972	0.3915	0.4176	0.4151	0.3987
MAE:	0.5187	0.5237	0.5164	0.5231	0.5226	0.5183	0.5304	0.5262	0.5296
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.6023	0.5871	0.5922	0.5833	0.6155	0.6254	0.5920	0.5956	0.5691
Std Y Pred:	0.0405	0.0412	0.0417	0.0416	0.0433	0.0413	0.0441	0.0491	0.0344
Avg Y Test:	0.6000	0.6354	0.6126	0.6161	0.6339	0.6362	0.6027	0.6196	0.6123
Std Y Test:	0.3806	0.4169	0.4002	0.3916	0.3818	0.4126	0.3946	0.3942	0.4177
R2 Score:	0.0018	0.0015	0.0086	-0.0031	0.0094	0.0094	0.0078	-0.0164	-0.0033
RMSE:	0.3802	0.4166	0.3985	0.3922	0.3800	0.4106	0.3931	0.3974	0.4184
MAE:	0.5172	0.5247	0.5225	0.5182	0.5177	0.5291	0.5213	0.5267	0.5290

Table 62. ERR@10 – Cluster 3

Cluster 4									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4279	0.4731	0.4202	0.4314	0.4800	0.4401	0.4456	0.4185	0.4608
Std Y Pred:	0.0469	0.0474	0.0464	0.0480	0.0538	0.0471	0.0468	0.0534	0.0349
Avg Y Test:	0.4742	0.5305	0.4512	0.4420	0.4868	0.4500	0.4562	0.4564	0.4938
Std Y Test:	0.3340	0.3835	0.3504	0.3320	0.3543	0.3354	0.3381	0.3593	0.3561
R2 Score:	-0.0057	-0.0102	0.0261	0.0278	0.0311	0.0226	0.0207	0.0008	0.0189
RMSE:	0.3350	0.3855	0.3458	0.3273	0.3487	0.3316	0.3346	0.3592	0.3528
MAE:	0.4900	0.5058	0.4905	0.4837	0.4923	0.4855	0.4842	0.4975	0.5015
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4353	0.4389	0.4294	0.4321	0.4856	0.4060	0.4448	0.4256	0.5052
Std Y Pred:	0.0487	0.0476	0.0453	0.0475	0.0571	0.0546	0.0498	0.0528	0.0402
Avg Y Test:	0.4678	0.4829	0.4520	0.4385	0.4933	0.4219	0.4387	0.4655	0.4779
Std Y Test:	0.3286	0.3572	0.3543	0.3321	0.3476	0.3454	0.3453	0.3332	0.3412
R2 Score:	0.0193	0.0097	0.0122	0.0123	0.0315	0.0420	0.0238	-0.0035	0.0099
RMSE:	0.3254	0.3554	0.3521	0.3300	0.3421	0.3381	0.3411	0.3338	0.3395
MAE:	0.4918	0.4936	0.4879	0.4871	0.4910	0.4872	0.4910	0.4874	0.4996
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.4505	0.4349	0.4393	0.4311	0.4599	0.4716	0.4262	0.4103	0.4497
Std Y Pred:	0.0478	0.0474	0.0451	0.0456	0.0489	0.0485	0.0532	0.0581	0.0357
Avg Y Test:	0.4705	0.4554	0.4559	0.4597	0.4527	0.4412	0.4028	0.4740	0.4860
Std Y Test:	0.3662	0.3541	0.3537	0.3425	0.3381	0.3345	0.3256	0.3590	0.3543
R2 Score:	0.0214	0.0213	0.0352	0.0146	0.0210	0.0319	0.0100	-0.0267	0.0121
RMSE:	0.3623	0.3503	0.3474	0.3400	0.3346	0.3292	0.3239	0.3638	0.3521
MAE:	0.5038	0.4888	0.4933	0.4892	0.4843	0.4886	0.4843	0.5035	0.4971

Table 63. ERR@10 – Cluster 4

Cluster 5									
SVM	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2550	0.3028	0.2389	0.2488	0.2625	0.2546	0.2674	0.2112	0.3325
Std Y Pred:	0.0601	0.0433	0.0635	0.0655	0.0791	0.0733	0.0632	0.0701	0.0443
Avg Y Test:	0.2506	0.3138	0.2237	0.2186	0.2454	0.2236	0.2418	0.2797	0.2685
Std Y Test:	0.2698	0.3071	0.2496	0.2502	0.2533	0.2464	0.2669	0.2755	0.2910
R2 Score:	0.0676	0.0681	0.0571	0.0346	0.1131	0.0389	0.0604	0.0323	0.0116
RMSE:	0.2605	0.2964	0.2424	0.2459	0.2385	0.2416	0.2588	0.2710	0.2893
MAE:	0.4362	0.4557	0.4251	0.4296	0.4228	0.4257	0.4403	0.4403	0.4652
LambdaMart	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2523	0.2573	0.2562	0.2469	0.2624	0.1918	0.2482	0.2268	0.3386
Std Y Pred:	0.0655	0.0634	0.0612	0.0653	0.0874	0.0818	0.0747	0.0695	0.0407
Avg Y Test:	0.2643	0.2677	0.2406	0.2098	0.2416	0.2134	0.2222	0.2913	0.2845
Std Y Test:	0.2670	0.2749	0.2699	0.2343	0.2440	0.2451	0.2477	0.2968	0.3009
R2 Score:	0.0759	0.0952	0.0779	0.0091	0.0886	0.0443	0.0475	0.0104	0.0186
RMSE:	0.2567	0.2615	0.2592	0.2333	0.2329	0.2396	0.2417	0.2953	0.2981
MAE:	0.4358	0.4339	0.4304	0.4178	0.4196	0.4079	0.4243	0.4539	0.4722
RankNet	Attention	Feedforward	GRU	GRU + Atten.	Light GBM	LSTM	LSTM + Atten.	Trans.	Trans. + MLP
Avg Y Pred:	0.2645	0.2531	0.2681	0.2572	0.2670	0.2787	0.2080	0.1876	0.3185
Std Y Pred:	0.0681	0.0742	0.0631	0.0664	0.0684	0.0813	0.0866	0.0745	0.0435
Avg Y Test:	0.2418	0.2422	0.2378	0.2427	0.2379	0.2214	0.2015	0.3342	0.2738
Std Y Test:	0.2406	0.2512	0.2506	0.2615	0.2440	0.2474	0.2363	0.3370	0.2661
R2 Score:	0.0583	0.0681	0.0565	0.0603	0.0757	0.0066	0.0472	-0.1139	0.0407
RMSE:	0.2335	0.2425	0.2434	0.2535	0.2346	0.2466	0.2306	0.3556	0.2606
MAE:	0.4236	0.4264	0.4317	0.4324	0.4289	0.4367	0.4073	0.4898	0.4527

Table 64. ERR@10 – Cluster 5

7 References

[1] Qingyao Ai, Keping Bi, Jiafeng Guo, and W Bruce Croft. 2018. Learning a deep listwise context model for ranking refinement. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. ACM, 135-- 144.

[2] Donald Metzler, Trevor Strohman, and W Bruce Croft. 2010. Search Engine Information Retrieval in Practice. Boston, MA: Pearson Education.

[3] Li, Anthony. (2016). Whom to Trust When Sick? The Wei Zexi Incident, the Chinese Internet, and the Healthcare System in China. China Perspectives. 2016. 79-83.

[4] Yajuan Duan, Long Jiang, Tao Qin, Ming Zhou, and Heung-Yeung Shum. 2010. An empirical study on learning to rank of tweets. In Proceedings of the 23rd Inter- national Conference on Computational Linguistics. Association for Computational Linguistics, 295–303.

[5] Tie-Yan Liu. 2009. Learning to rank for information retrieval. Foundations and Trends in Information Retrieval 3, 3 (2009), 225–331.

[6] Y. Zhou and B. W. Croft. Query performance prediction in web search environments. In phSIGIR '07, 2007.

[7] R. Herbrich, K. Obermayer, and T. Graepel. Large margin rank boundaries for ordinal regression. In Advances in Large Margin Classifiers, pages 115–132, Cambridge, MA, 1999. MIT.

[8] P. Li, C. Burges, and Q. Wu. Mcrank: Learning to rank using multiple classification and gradient boosting. In NIPS ’07: Advances in Neural Information Processing Systems 20, pages 897–904, Cambridge, MA, 2008. MIT.

[9] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. Learning to rank: from pairwise approach to listwise approach. In ICML ’07: Proceedings of the 24th International Conference on Machine learning, pages 129–136, New York, NY, USA, 2007. ACM.

[10] F. Xia, T.-Y. Liu, J. Wang, W. Zhang, and H. Li. Listwise approach to learning to rank - theory and algorithm. In ICML ’08: Proceedings of the 25th International Conference on Machine learning, pages 1192–1199. Omnipress, 2008.

[11] Yining Wang, Liwei Wang, Yuanzhi Li, Di He, Wei Chen, Tie-Yan Liu. 2013. A Theoretical Analysis of Normalized Discounted Cumulative Gain (NDCG) Ranking Measures. In Proceedings of the 26th Annual Conference on Learning Theory (COLT 2013).

[12] Shtok, A., Kurland, O., and Carmel, D. 2010. Using statistical decision theory and relevance models for query performance prediction. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. 259-

[13] Y. Duan, L. Jiang, T. Qin, M. Zhou, and H.-Y. Shum. An empirical study on learning to rank of tweets. In COLING, 2010.

[14] T. Y. Liu. Learning to rank for information retrieval. Foundations and Trends in Information Retrieval, 3(3), pp. 23–24, 2009.

[15] B. He and I. Ounis. Query performance prediction. Information Systems, 31:585–594, 2006.

[16] S. Cronen-Townsend, Y. Zhou, and W. B. Croft. Predicting query performance. In 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 299–306, 2002.

[17] S. Robertson and H. Zaragoza. The probabilistic relevance framework: BM25 and beyond. Foundations and Trends in Information Retrieval, 3(4):333–389, 2009.

[18] S. E. Robertson and D. A. Hull. The TREC-9 filtering track final report. In TREC, pages 25–40, 2000

[19] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to Ad Hoc information retrieval. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, pages 334–342. ACM Press New York, NY, USA, 2001.

[20] O. Chapelle, D. Metlzer, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In CIKM ’09: Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009.

[21] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In ICML ’05: Proceedings of the 22nd international conference on Machine learning, pages 89–96, New York, NY, USA, 2005. ACM Press.

[22] Yanyan Lan, Yadong Zhu, Jiafeng Guo, Shuzi Niu, and Xueqi Cheng. 2014. Position-Aware ListMLE: a Sequential Learning Process for Ranking. In Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence. 449–458.

[23] O. Chapelle and Y. Chang. Yahoo! Learning to Rank Challenge Overview. Journal of Machine Learning Research - W & CP, 14:1–24, 2011.

[24] Severyn, Aliaksei and Moschitti, Alessandro. Learning to rank short text pairs with convolutional deep neural networks. Proceedings of SIGIR, 2015.

[25] A. F. Agarap. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375, 2018.

[26] L. Lu, Y. Shin, Y. Su, and G. E. Karniadakis, Dying ReLU and Initialization: Theory and Numerical Examples, preprint, https://arxiv.org/abs/1903.06733, 2019.

[27] Xingyuan Pan and Vivek Srikumar. Expressiveness of rectifier networks. In ICML, pages 2427–2435, 2016.

[28] D. Cohen, J. Foley, H. Zamani, J. Allan, W. B. Croft, Universal approximation functions for fast learning to rank: Replacing expensive regression forests with simple feed-forward networks, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '18, ACM, New York, NY, USA, 2018, pp. 1017–1020.

[29] Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neural network. *Chemometr Intell Lab Syst* 1997; 39(1): 43–62.

[30] Cesar Laurent, Gabriel Pereyra, Phil ´ emon Brakel, Ying Zhang, and Yoshua Bengio. Batch normalized recurrent ´ neural networks. arXiv preprint arXiv:1510.01378, 2015.

[31] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[32] Toharudin, Toni & Pontoh, Resa & Caraka, Rezzy & Zahroh, Solichatus & Lee, Youngjo & Chen, Rung. (2021). Employing Long Short-Term Memory and Facebook Prophet Model in Air Temperature Forecasting. Communication in Statistics- Simulation and Computation. 10.1080/03610918.2020.1854302.

[33] Faghfouri, Aram & Frish, Michael. (2011). Robust discrimination of human footsteps using seismic signals. Proc SPIE. 10.1117/12.882726.

[34] Junyoung Chung, Çaglar Gülçehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. CoRR, abs/1412.3555, 2014

[35] Jing, R. (2019). A Self-attention Based LSTM Network for Text Classification, Journal of Physics: Conference Series 1207(1).

[36] Kim, S., & Kang, M. (2019). Financial series prediction using attention lstm. arXiv preprint arXiv:1902.10877 , .

[37] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to Attention-based neural machine translation. arXiv preprint arXiv:1508.04025, 2015.

[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. arXiv preprint arXiv:1706.03762, 2017.

[39] Thorsten Joachims, *Learning to Classify Text Using Support Vector Machines*. Dissertation, Kluwer, 2002.

[40] Dang, V. "The Lemur Project-Wiki-RankLib." Lemur Project, [Online]. Available: http://sourceforge. net/p/lemur/wiki/RankLib.

[41] Zhen Qin, Le Yan, Honglei Zhuang, Yi Tay, Rama Kumar Pasumarthi, Xuanhui Wang, Michael Bendersky, and Marc Najork. 2021. Are Neural Rankers Still Outperformed by Gradient Boosted Decision Trees? In Proceedings of the 9th International Conference on Learning Representations (ICLR '21).

[42] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. arXiv preprint arXiv:1603.06560, 2016.

[43] user124589 (2016, Oct 23). *Understanding LSTM units vs. Cells* [Discussion post]. Stack Exchange Network https://stats.stackexchange.com/questions/241985/understanding-lstm-units-vs-cells

[44] Seunghyun Yoon, Joongbo Shin, and Kyomin Jung. Learning to rank question-answer pairs using hierarchical recurrent encoder with latent topic clustering. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 1575–1584,

New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1142.

[45] Lee, H.C., Rim, H.C. and Lee, D.G., 2019. Learning to rank products based on online product reviews using a hierarchical deep neural network. Electronic Commerce Research and Applications, 36, p.100874.

[46] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems, pages 3149–3157, 2017.