

--	--	--	--	--	--	--	--

## KALASALINGAM ACADEMY OF RESEARCH AND EDUCATION

(Deemed to be University)

Anand Nagar, Krishnankoil – 626 126.

### END SEMESTER EXAMINATIONS - NOV/DEC 2024

Course Code	:	224MDS3115	Duration	:	180 Minutes
Course Name	:	Recommendation System	Max. Marks	:	100
Degree	:	M.Sc.			

<b>PART – A (20 Marks) Answer All Questions</b>				<b>Marks</b>
1	Mention about Netflix recommender system			2.0
	<b>Key:</b> <p>Netflix provides users the ability to rate the movies and television shows on a 5-point scale. Furthermore, the user actions in terms of watching various items are also stored by Netflix. These ratings and actions are then used by Netflix to make recommendations. Netflix does an excellent job of providing explanations for the recommended items.</p>			
2	How user-based collaborative filtering works?			2.0
	<b>Key:</b> <p>User-based collaborative filtering uses the ratings of top-k similar users in predicting the rating of a specific item for the target user</p>			
3	Name the most desired rating system suitable for building recommenders using association analysis?			2.0
	<b>Key:</b> <p>The most suitable rating system suitable for building recommenders using association analysis is unary rating systems. Unary rating systems indicate whether an user has made a purchase or not. Other machine learning methods lack the ability to handle such data sets.</p>			
4	Mention how the problem of sparseness is dealt with while building decision trees for recommenders.			2.0
	<b>Key:</b> <p>The problem of sparseness can be handled by using a dimension reduction technique. A reduced dimensional space creates dense data set.</p>			
5	What are the two types of data used in content based recommendation systems?			2.0
	<b>Key:</b> <p>1. The first source of data is a description of various items in terms of content-centric attributes. An example of such a representation could be the text description of an item by the manufacturer. 2. The second source of data is a user profile, which is generated from user feedback about various items. The user feedback might be explicit or implicit. Explicit feedback may correspond to ratings, whereas implicit feedback may correspond to useractions.</p>			

<b>PART – A (20 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
6	What are damping functions?  <b>Key:</b> Excessive occurrence of a single word should not be given too much importance. This is likely to happen when item descriptions are collected from unreliable sources or open platforms, such as the Web. To achieve this goal, a damping function is optionally applied to the frequencies before similarity computation.	2.0
7	Give an example of a symmetric metric used in case based recommendation systems  <b>Key:</b> Consider the attribute $i$ associated with an item having several attributes. Denote by, $t_i$ its value with respect to target item and $x_i$ be the corresponding value with respect to the item being considered. Then, the similarity between the two items with respect to this attribute is defined as follows: $\text{Sim}(t_i, x_i) = 1 - [ \{t_i - x_i\}  / \{\max_i - \min_i\}]$	2.0
8	Give example for direct mapping in knowledge bases  <b>Key:</b> These rules relate customer requirements to hard requirements on product attributes. An example of such a rule is as follows: Suburban-or-rural=Suburban $\Rightarrow$ Locality=(List of relevant localities) Min-Bedrooms $\geq 3 \Rightarrow$ Price $\geq \$ 100,000$	2.0
9	Why it is necessary to consider experimental design issues while developing recommender evaluation systems?  <b>Key:</b> It is crucial to design the experiments so that the accuracy is not overestimated or underestimated. For example, if the same set of specified ratings is used both for model construction and for accuracy evaluation, then the accuracy will be grossly overestimated. In this context, careful experimental design is important	2.0
10	Name the quantities used in defining rating-based utility of an item  <b>Key:</b> It uses the user's rating of the item, a predefined break-even value to define rating-based utility of an item	2.0

**PART – B (80 Marks)**  
**Answer All Questions**

**Marks**

- 11.A Given the mean centered-rating matrix as given below.

	i1	i2	i3	i4	i5	i6	Means
u1	1.5	0.5	1.5	-1.5	-0.5	-1.5	5.5
u2	1.2	2.2?		-0.8	-1.8	-0.8	4.8
u3?		1	1	-1	-1?		2
u4	-1.5	-0.5	-0.5	0.5	0.5	1.5	2.5
u5	-1		-1	0	1	1	2

Estimate the rating for user 3 item 1 using UBCF using k=2

**Key:**

**PART – B (80 Marks)**  
**Answer All Questions**

**Marks**

The Pearson-similarity values are obtained as follows:

sim(3,1)		sim(3,2)		sim(3,4)		(3,5)	
x	y	x	y	x	y	x	y
1	0.5	1	2.2	1	-0.5	1	-1
1	1.5	-1	-0.8	1	-0.5	-1	0
-1	-1.5	-1	-1.8	-1	0.5	-1	1
-1	-0.5			-1	0.5		
	0.8944		0.9707		-1		0.8660

Top-2 similar users are 2 and 1 with similarity values 0.9707 and 0.8944.

Their mean centered ratings for the item-1 are 1.2 and 1.5 respectively.

Hence, estimated rating is

$$2 + \left( \frac{0.970725 \cdot 1.2 + 0.894427 \cdot 1.5}{0.970725 + 0.894427} \right) = 3.3438$$

13

OR

**PART – B (80 Marks)**  
**Answer All Questions**

**Marks**

11.B

16.0

i1	i2	i3	i4	i5	i6
7	6	7	4	5	4
6	7	?	4	3	4
?	3	3	1	1	?
1	2	2	3	3	4
1	?	1	2	3	3

Estimate the rating  $r_{36}$  using item-based collaborative filtering using cosine similarity assuming  $k=2$

**Key:**

i6	i1	i6*i1	i6^2	i1^2	i6	i2	i6*i2	i6^2	i2^2	i6	i5	i6^2	i6*i6	i5^2
4	7	28	16	49	4	6	24	16	36	4	5	20	16	
4	7	28	16	49	4	7	28	16	49	4	3	12	16	
4	1	4	16	1	4	2	8	16	4	4	3	12	16	
3	1	3	9	1			60	48	89	3	3	9	9	
		63	57	100		12	15					53	57	
		r_61	0.8			r_62	0.918					r_65	0.97	

i6	i3	i6*i3	i6^2	i3^2	i6	i4	i6*i4	i6^2	i4^2
4	7	28	16	49	4	4	16	16	16
4	2	8	16	4	4	4	16	16	16
3	1	3	9	1	4	3	12	16	9
		r_63	0.8		3	2	6	9	4
							50	57	45
							r_64	0.987	

Most 2-similar items are 4 and 5 with sim values 0.987 and 0.97;

Their ratings by the target user are 4 and 3

Hence, the predicted rating is 4.496

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
12.A	<p>Write a descriptive note on Model Based Collaborative Filtering Systems</p> <p><b>Key:</b></p> <p>In model-based methods, a summarized model of the data is using supervised or unsupervised machine learning methods. Therefore, the training (or model building phase) is clearly separated from the prediction phase. Examples of such methods in traditional machine learning include decision trees, rule-based methods, Bayes classifiers, regression models, support vector machines, and neural networks . Association analysis is used for building recommenders using association analysis in unary rating systems. Unary rating systems indicate whether an user has made a purchase or not. Other machine learning methods lack the ability to handle such data sets. Naive Bayesian recommender systems are built if (1) The number of possible ratings is small each of which can be treated as a categorical value and (2) Orderings among the ratings are ignored. Regression trees are used while building model-based collaborative filtering when the rating to be predicted is interval-valued. Classification trees are used while building model-based collaborative filtering when the rating to be predicted is of categorical type.</p> <p>Advantages :</p> <p><b>Space Efficiency</b> Typically, the size of the learned model is much smaller than the original ratings matrix. Thus, the space requirements are often quite low. On the other hand, a user-based neighborhood method might have <math>O(m^2)</math> space complexity, where m is the number of users. An item-based method will have <math>\\$ O(n^2) \\$</math> space complexity.</p> <p><b>Speed</b> Model-based recommender systems often have a number of advantages over neighborhood-based methods: Model-based systems are usually much faster in the preprocessing phase of constructing the trained model. In most cases, the compact and summarized model can be used to make predictions efficiently.</p> <p><b>Avoiding Over fitting</b> Overfitting is a serious problem in many machine learning algorithms, in which the prediction is overly influenced by random artifacts in the data. This problem is also encountered in classification and regression models. The summarization approach of model-based methods can often help in avoiding overfitting. Furthermore, regularization methods can be used to make these models robust.</p>	16.0

OR

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
12.B	<p>Consider a rating matrix with 8 users and items with possible ratings 1,2 and 3 as given below</p> $\begin{bmatrix} & i1 & i2 & i3 & i4 \\ u1 & 1 & 2 & 1 & 3 \\ u2 & 2 & 2 & 3 & 1 \\ u3 & 3 & 2 & 1 & 3 \\ u4 & 1 & 2 & 1 & 3 \\ u5 & ? & 2 & 1 & 3 \\ u6 & 2 & 1 & 1 & 3 \\ u7 & 1 & 2 & 1 & 3 \\ u8 & 1 & 3 & 2 & 1 \end{bmatrix}$ <p>Compute the predicted rating of item 1 for user 5 using the two approaches of naïve Bayesian collaborative filtering</p>	16.0

**Key:**

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>	<b>Marks</b>
$P(r_{51} = 1 i_2 = 2, i_3 = 1, i_4 = 3)$ $\propto P(r_{51} = 1)P(i_2 = 2 r_{51} = 1)P(i_3 = 1 r_{51} = 1)P(i_4 = 3 r_{51} = 1)$ $= \frac{4}{7} \frac{3}{4} \frac{3}{4} \frac{3}{4} = \frac{27}{112}$ $P(r_{51} = 2 i_2 = 2, i_3 = 1, i_4 = 3)$ $\propto P(r_{51} = 2)P(i_2 = 2 r_{51} = 2)P(i_3 = 1 r_{51} = 2)P(i_4 = 3 r_{51} = 2)$ $= \frac{2}{7} \frac{1}{2} \frac{1}{2} \frac{1}{2} = \frac{1}{28}$ $P(r_{51} = 3 i_2 = 2, i_3 = 1, i_4 = 3)$ $\propto P(r_{51} = 3)P(i_2 = 2 r_{51} = 3)P(i_3 = 1 r_{51} = 3)P(i_4 = 3 r_{51} = 3)$ $= \frac{1}{7} \frac{1}{1} \frac{1}{1} \frac{1}{1} = \frac{1}{7}$ <p>Thus, we have</p> $P(r_{51} = 1 i_2 = 2, i_3 = 1, i_4 = 3) = k \frac{27}{112}$ $P(r_{51} = 2 i_2 = 2, i_3 = 1, i_4 = 3) = k \frac{4}{112}$ $P(r_{51} = 3 i_2 = 2, i_3 = 1, i_4 = 3) = k \frac{16}{112}$ <p>Hence, by Approach-1, the predicted rating is <math>\hat{r}_{51} = 3</math></p> <p><b>Approach-2</b> Note that, the value of <math>k</math> is obtained by solving the equation,</p> $k \frac{27}{112} + k \frac{4}{112} + k \frac{16}{112} = 1$ <p>Hence, <math>k = \frac{112}{47}</math></p> <p>Therefore, by Approach-2, <math>\hat{r}_{51} = 1 \times \frac{27}{47} + 2 \times \frac{4}{47} + 3 \times \frac{16}{47} = 1.76 \approx 2</math></p>	

PART – B (80 Marks)						Marks																																					
Answer All Questions																																											
13.A	<p>Consider a hypothetical situation involving a rating system with 4 possible ratings namely 1, 2, 3 and 4.</p> <p>Assume there are 5 keywords and 100 documents.</p> <p>The data given below explains distributions of keywords over various documents along with ratings.</p> <table border="1"> <tr> <td></td><td>1</td><td>2</td><td>3</td><td>4</td><td></td></tr> <tr> <td>k1</td><td>8</td><td>5</td><td>0</td><td>3</td><td></td></tr> <tr> <td>k2</td><td>12</td><td>10</td><td>5</td><td>2</td><td></td></tr> <tr> <td>k3</td><td>10</td><td>5</td><td>10</td><td>4</td><td></td></tr> <tr> <td>k4</td><td>5</td><td>6</td><td>5</td><td>1</td><td></td></tr> <tr> <td>k5</td><td>5</td><td>4</td><td>0</td><td>0</td><td></td></tr> </table>							1	2	3	4		k1	8	5	0	3		k2	12	10	5	2		k3	10	5	10	4		k4	5	6	5	1		k5	5	4	0	0		16.0
	1	2	3	4																																							
k1	8	5	0	3																																							
k2	12	10	5	2																																							
k3	10	5	10	4																																							
k4	5	6	5	1																																							
k5	5	4	0	0																																							

Using entropies, identify the most discriminative and least discriminative features.

**Key:**

Entropy for keywords are given below:

$$e(k_1) = -[(\log \frac{8}{16}) \frac{8}{16} + (\log \frac{5}{16}) \frac{5}{16} + (\log \frac{3}{16}) \frac{3}{16}] = 0.44$$

$$e(k_2) = -[(\log \frac{12}{29}) \frac{12}{29} + (\log \frac{10}{29}) \frac{10}{29} + (\log \frac{5}{29}) \frac{5}{29} + (\log \frac{2}{29}) \frac{2}{29}] = 0.53$$

$$e(k_3) = -[(\log \frac{10}{29}) \frac{10}{29} + (\log \frac{5}{29}) \frac{5}{29} + (\log \frac{10}{29}) \frac{10}{29} + (\log \frac{4}{29}) \frac{4}{29}] = 0.569$$

$$e(k_4) = -[(\log \frac{5}{17}) \frac{5}{17} + (\log \frac{6}{17}) \frac{6}{17} + (\log \frac{5}{17}) \frac{5}{17} + (\log \frac{1}{17}) \frac{1}{17}] = 0.541$$

$$e(k_5) = -[(\log \frac{5}{9}) \frac{5}{9} + (\log \frac{4}{9}) \frac{4}{9}] = 0.298$$

The keyword corresponding to smallest entropy must be taken as most discriminative. In this case it happens to be  $k_5$ .

OR

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
13.B	<p>Describe in detail the general framework of Learning User Profiles and Filtering</p> <p><b>Key:</b></p> <p>The learning of user profiles is closely related to the classification and regression modeling problem.</p> <p>When the ratings are treated as discrete values (e.g., “thumbs up” or “thumbs down”), the problem is similar to that of text classification.</p> <p>On the other hand, when the ratings are treated as a set of numerical entities, the problem is similar to that of regression modeling.</p> <ul style="list-style-type: none"> <li>1. Assume that <math>DL</math> be the set of training documents, which are labeled by a specific user</li> <li>2. This user is also referred to as the active user when that user obtains a recommendation from the system.</li> <li>3. The training documents correspond to the descriptions of items, which are extracted in the preprocessing and feature selection phases.</li> <li>4. Furthermore, the training data contain the ratings assigned by the active user to these documents.</li> <li>5. These documents are used to construct a training model.</li> <li>6. Note that the labels assigned by other users are not used in the training process.</li> <li>7. Therefore, the training models are specific to particular users, and they cannot be used for arbitrarily chosen users</li> <li>8. The training model for a specific user represents the user profile.</li> <li>9. The labels on the documents correspond to the numeric, binary, or unary ratings</li> <li>10. Assume that the <math>i</math>th document in <math>DL</math> has a rating denoted by <math>c_i</math>. We also have a set <math>\{D\}_U</math> of testing documents, which are unlabeled. The testing documents might correspond to descriptions of items, which might be potentially recommended to the user but which have not yet been bought or rated by the user. In domains such as news recommendation the documents in <math>\{D\}_U</math> might correspond to candidate Web documents for recommendation to the active user. The precise definition of <math>\{D\}_U</math> depends on the domain at hand, but the individual documents in <math>\{D\}_U</math> are extracted in a similar way to those in <math>\{D\}_L</math>. The training model on <math>\{D\}_L</math> is used to make recommendations from <math>\{D\}_U</math> to the active user. As in the case of collaborative filtering, the model can be used to provide either a predicted value of the rating or a ranked list of top-<math>k</math> recommendations. It is obvious that this problem is similar to that of classification and regression modeling in the text domain.</li> </ul>	16.0

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
14.A	<p>(i) Explain the three phases in which the interaction between user and recommender system generally proceeds</p> <p><b>Key:</b></p> <p>In the first phase, an interactive interface is used by the user to specify his initial preferences. A common approach is to use a Web style form in which the desired values of the attributes may be entered. For example, the car recommendation site Edmunds.com presents a series of interfaces to the users to specify their preferences about the specific features they might want. The answers to the queries in the first interface may affect the questions in the next interface.</p> <p>In the second phase, the user is presented with a ranked list of matching items. An explanation for why the items are returned is typically provided. In some cases, no items might match the user requirements. In such cases, possible relaxations of the requirements might be suggested. In cases, where too many items are returned, suggestions for possible constraints (user requirements) are included.</p> <p>In the third phase, the user refines his requirements depending on the returned results. This refinement might take the form of the addition of further requirements, or the removal of some of the requirements. For example, when an empty set is returned, it is evident that some of the requirements need to be relaxed. Constraint satisfaction methods are used to identify possible sets of candidate constraints, which might need to be relaxed. Therefore, the system generally helps the user in making her modifications in a more intelligent and efficient way.</p>	8.0

PART – B (80 Marks) Answer All Questions		Marks																
(ii) Give an example user interface for dynamic critiquing in a case-based recommender	8.0																	
<p><b>Key:</b></p> <p>The Figure given below is an example user interface associated with dynamic critiquing in case based recommendations.</p> <p><b>EXAMPLE OF HYPOTHETICAL CASE-BASED RECOMMENDATION INTERFACE FOR HOME BUYING (critique-example.com)</b></p> <p>[ DYNAMIC CRITIQUING INTERFACE ]</p> <p>YOU SPECIFIED THE FOLLOWING TARGET: 812 SCENIC DRIVE, MOHEGAN LAKE, NY</p> <p>YOUR TOP RECOMMENDATION IS: 742 SCENIC DRIVE, MOHEGAN LAKE, NY</p> <p>WE RECOMMEND THIS HOUSE BECAUSE: IT HAS SIMILAR BEDROOMS, BATHROOMS, LOCALITY, PRICE RANGE, AND HOME STYLE AS YOUR TARGET</p> <p>I WOULD LIKE TO BUY A HOUSE SIMILAR TO THE TOP RECOMMENDATION BUT WITH ONE OF THE FOLLOWING CHANGE COMBINATIONS :</p> <table border="0"> <tr> <td>DIFFERENT STYLE AT SMALLER PRICE (12)</td> <td>SUBMIT CHANGE</td> <td>MORE BEDROOMS AT GREATER PRICE (22)</td> <td>SUBMIT CHANGE</td> </tr> <tr> <td>FEWER BEDROOMS AT SMALLER PRICE (13)</td> <td>SUBMIT CHANGE</td> <td>DIFFERENT STYLE IN NEARBY LOCALITY (29)</td> <td>SUBMIT CHANGE</td> </tr> <tr> <td>MORE BEDROOMS IN NEARBY LOCALITY (15)</td> <td>SUBMIT CHANGE</td> <td colspan="2" style="text-align: center;">SEE OTHER RESULTS</td> </tr> <tr> <td></td> <td></td> <td colspan="2" style="text-align: center;">GO BACK TO ENTRY POINT</td> </tr> </table> <p>OR</p>			DIFFERENT STYLE AT SMALLER PRICE (12)	SUBMIT CHANGE	MORE BEDROOMS AT GREATER PRICE (22)	SUBMIT CHANGE	FEWER BEDROOMS AT SMALLER PRICE (13)	SUBMIT CHANGE	DIFFERENT STYLE IN NEARBY LOCALITY (29)	SUBMIT CHANGE	MORE BEDROOMS IN NEARBY LOCALITY (15)	SUBMIT CHANGE	SEE OTHER RESULTS				GO BACK TO ENTRY POINT	
DIFFERENT STYLE AT SMALLER PRICE (12)	SUBMIT CHANGE	MORE BEDROOMS AT GREATER PRICE (22)	SUBMIT CHANGE															
FEWER BEDROOMS AT SMALLER PRICE (13)	SUBMIT CHANGE	DIFFERENT STYLE IN NEARBY LOCALITY (29)	SUBMIT CHANGE															
MORE BEDROOMS IN NEARBY LOCALITY (15)	SUBMIT CHANGE	SEE OTHER RESULTS																
		GO BACK TO ENTRY POINT																

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
14.B	<p>(i) Explain how utility functions are used to rank matched items in constrained based recommenders</p> <p><b>Key:</b></p> <p>Ans : A common approach is to use utility functions in order to rank the matched items. Let <math>V = (v_1 \dots v_d)</math> be the vector of values defining the attributes of the matched products. Therefore, the dimensionality of the content space is <math>d</math>. The utility functions may be defined as weighted functions of the utilities of individual attributes. Each attribute has a weight <math>w_j</math> assigned to it, and it has a contribution defined by the function <math>f_j(v_j)</math> depending on the value <math>v_j</math> of the matched attribute. Then, the utility <math>U(V)</math> of the matched item is given by the following:</p> $U(V) = \sum_{j=1}^d w_j f_j(v_j)$ <p>Clearly, one needs to instantiate the values of <math>w_j</math> and <math>f_j(\cdot)</math> in order to learn the utility function. The design of effective utility functions often requires domain-specific knowledge, or learning data from past user interactions. For example, when <math>v_j</math> is numeric, one might assume that the function <math>f_j(v_j)</math> is linear in <math>v_j</math>, and then learn the coefficients of the linear function as well as <math>w_j</math> by eliciting feedback from various users. Typically, training data is obtained from some users who are given the task of ranking some sample items. These ranks are then used to learn the above mentioned model with the use of regression models.</p>	8.0

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
	<p>(ii) Explain the three forms of interaction between user and recommender.</p> <p><b>Key:</b></p> <p>Conversational systems: In this case, the user preferences are determined in the context of a feedback loop. The main reason for this is that the item domain is complex, and the user preferences can be determined only in the context of an iterative conversational system</p> <p>Search-based systems: In search-based systems, user preferences are elicited by using a preset sequence of questions such as the following: “Do you prefer a house in a suburban area or within the city?”</p> <p>Navigation-based recommendation: In navigation-based recommendation, the user specifies a number of change requests to the item being currently recommended. Through an iterative set of change requests, it is possible to arrive at a desirable item. An example of a change request specified by the user, when a specific house is being recommended is as follows: “I would like a similar house about 5 miles west of the currently recommended house.” Such recommender systems are also referred to as critiquing recommender systems</p>	8.0

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
15.A	<p>(i) Explain : Spearman Correlation and Kendall's tau  (ii) . Consider a user John, for whom you have hidden his ratings for Aliens (5), Terminator (5), Nero (1), and Gladiator (6).  The values in brackets represent his hidden ratings, and higher values are better.  Now consider a scenario where the recommender system ranks these movies in the order Terminator, Aliens, Gladiator, Nero.</p> <p>a. Compute the Spearman rank correlation coefficient as a measure of recommendation ranking quality  b. Compute the Kendall rank correlation coefficient as a measure of ranking quality.}</p> <p><b>Key:</b></p>	16.0

**PART – B (80 Marks)**  
**Answer All Questions**

**Marks**

- Spearman correlation: The formula for Spearman correlation when there are no ties is given by  $1 - 6 \frac{\sum_{i=1}^n d_i^2}{n(n^2-1)}$  where  $d_i$  is the difference between the actual and predicted ranks for the item  $i$ . When there are ties, the rank correlation is computed using the formula

$$1 - 6 \frac{\sum_{i=1}^n d_i^2 + \frac{1}{12} m_1(m_1^2 - 1) + \frac{1}{12} m_2(m_2^2 - 1)}{n(n^2 - 1) + \dots}$$

$m_i$  indicate number of tied items for rank  $i$

- Kendall correlation is defined by  $\tau = 2 \frac{n_c - n_d}{n(n-1)}$  where  $n_c$  is the number of concordant pairs and  $n_d$  number of discordant pairs.

The following table shows calculations related to Spearman-correlation.

Movie	Rating	Ranking(Actual)	Ranking (Est)	$d_i$	$d_i^2$
Aliens	5	2.5	2	-0.5	0.25
Terminator	5	2.5	1	1.5	2.25
Nero	1	4	4	0	0
Gladiator	6	1	3	2	4
					6.5

There are two items with tied ranks. Hence, the Spearman correlation is

$$1 - 6 \frac{\sum_{i=1}^n d_i^2 + \frac{1}{12} m_1(m_1^2 - 1) + \frac{1}{12} m_2(m_2^2 - 1)}{n(n^2 - 1) + \dots} = 1 - 6 \frac{6.5 + \frac{1}{12}(2)(2^2 - 1)}{4(4^2 - 1)} = 1 - \frac{42}{60} = 0.3$$

Pairs of Movies	Ranking pairs	Concordance/Discordance
Aliens, Terminator	(2.5,2.5),(2,1)	-
Aliens, Nero	(2.5,4),(2,4)	+
Aliens, Gladiator	(2.5,1),(2,3)	-
Terminator, Nero	(2.5,4),(1,4)	+
Terminator, Gladiator	(2.5,1),(1,3)	-
Nero, Gladiator	(4,1),(4,3)	+

Here ,  $n_c = 3$  and  $n_d = 6$ . Hence, Kendall correlation is

$$\tau = 2 \frac{n_c - n_d}{n(n-1)} = \frac{3-3}{6} = 0$$

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>		<b>Marks</b>
15.B	Explain any five measures meant for measuring the accuracy of ratings predictions  <b>Key:</b>	16.0

<b>PART – B (80 Marks)</b> <b>Answer All Questions</b>	<b>Marks</b>
<p><b>Ans :</b> The entry-specific error of the estimation is given by the quantity <math>e_{uj} = \hat{r}_{uj} - r_{uj}</math> for user u and item j. This error can be used in various ways to compute the overall error over the set E of entries in the ratings matrix on which the evaluation is performed. An example is the mean squared error, which is denoted by MSE:</p> $MSE = \frac{\sum_{(u,j) \in E} e_{uj}^2}{ E }$ <p>. The square-root of the aforementioned quantity is referred to as the root mean squared error, or</p> $RMSE = \sqrt{\frac{\sum_{(u,j) \in E} e_{uj}^2}{ E }}$ <p>The RMSE is in units of ratings, rather than in units of squared ratings like the MSE. One characteristic of the RMSE is that it tends to disproportionately penalize large errors because of the squared term within the summation. One measure, known as the mean-absolute-error (MAE), does not disproportionately penalize larger errors.</p> $MAE = \frac{\sum_{(u,j) \in E}  e_{uj} }{ E }$ <p>Other related measures such as the normalized RMSE (NRMSE) and normalized MAE (NMAE) are defined in a similar way, except that each of them is divided by the range <math>r_{max} - r_{min}</math> of the ratings:</p> $NRMSE = \frac{MSE}{r_{max} - r_{min}}$ $NMAE = \frac{MAE}{r_{max} - r_{min}}$ <p>The normalized values of the RMSE and MAE always lie in the range (0, 1), and therefore they are more interpretable from an intuitive point of view. It is also possible to use these values to compare the performance of a particular algorithm over different data sets with varying scales of ratings.</p>	

\*\*\*\*\*