**INFORMATION TECHNOLOGY AND MANAGEMENT**

**REVISED HOMEWORK #3 – Fall 2020**

**Due November 17, 2020**

1. **Using the following relations and subordinate dependencies convert these relations into an NER diagram. (1) Draw the subordinate dependency graph, (2) “circle” RHMOs, (3) draw NER diagram, and (4) repeat these steps for each complex object at the next (more detailed) level of abstraction.**

**CORP-UNITS ( UNIT-ID, CORP-ID, V-PRES )**

**SALES-DIVISION (SALES-DIV#, UNIT-ID)**

**EMPLOYEES ( EMP#, UNIT-ID, JOB-TITLE)**

**WAREHOUSE-EMP ( W-EMP#, WAREHOUSE-ID )**

**WAREHOUSES (WAREHOUSE-ID, WAREHOUSE-SITE,**

**WRHSE-DIR# )**

**WAREHOUSE-ITEM-LOC ( WAREHOUSE-ID, ITEM-ID, AISLE#, BIN#, QNTY )**

**DELIVERY-ORDER ( D-ORD#, CUST-ORDER#, W-EMP# )**

**INV-ITEMS ( ITEM-ID, ITEM-DESCR, UNIT-PRICE )**

**SALES-EMP ( S-EMP#, SALES-DIV# )**

**SUP-ORDERS ( SUP-ORDER#, DATE, SUPPLIER-ID )**

**SUP-ORDER-ITEMS ( SUP-ORDER#, ITEM-ID, QNTY-ORDERED )**

**SUPPLIER ( SUPPLIER-ID, SALES-CONTACT )**

**SUPPLIER-STOCK ( SUPPLIER-ID, ITEM-ID, UNIT-PRICE )**

**CUSTOMERS ( CUST-ID, CUST-SITE, CUST-ACCT )**

**CUST-ORDERS ( CUST-ORDER#, CUST-ID, S-EMP#, DEL-DATE )**

**CUST-ORDER-ITEMS ( CUST-ORDER#, ITEM-ID, QNTY-ORDERED )**

**SALES-DIVISION.UNIT-ID 🡪 CORP-UNIT.UNIT-ID**

**EMPLOYEES.UNIT-ID 🡪 CORP-UNIT.UNIT-ID**

**WAREHOUSE-EMP.W-EMP# 🡪 EMPLOYEES.EMP#**

**WAREHOUSE-EMP.WAREHOUSE-ID 🡪 WAREHOUSES.WAREHOUSE-ID**

**WAREHOUSE-ITEM-LOC.ITEM-ID 🡪 INV-ITEMS.ITEM-ID**

**WAREHOUSE-ITEM.LOC.WAREHOUSE-ID 🡪 WAREHOUSES.WAREHOUSE-ID**

**DELIVERY-ORDER.CUST-ORDER# 🡪 CUST-ORDERS.CUST-ORDER#**

**DELIVERY-W-EMP# 🡪 WAREHOUSE-EMP.W-EMP#**

**SALES-EMP.S-EMP# 🡪 EMPLOYEES.EMP#**

**SALES-EMP.SALES-DIV# 🡪 SALES-DIVISION.SALES-DIV#**

**SUP-ORDERS.SUPPLIER-ID 🡪 SUPPLIER.SUPPLIER-ID**

**SUP-ORDER-ITEMS.SUP-ORDER# 🡪 SUP-ORDERS.SUP-ORDER#**

**SUP-ORDER-ITEMS.ITEM-ID 🡪 INV-ITEMS.ITEM-ID**

**SUPPLIER-STOCK.SUPPLIER-ID 🡪 SUPPLIER.SUPPLIER-ID**

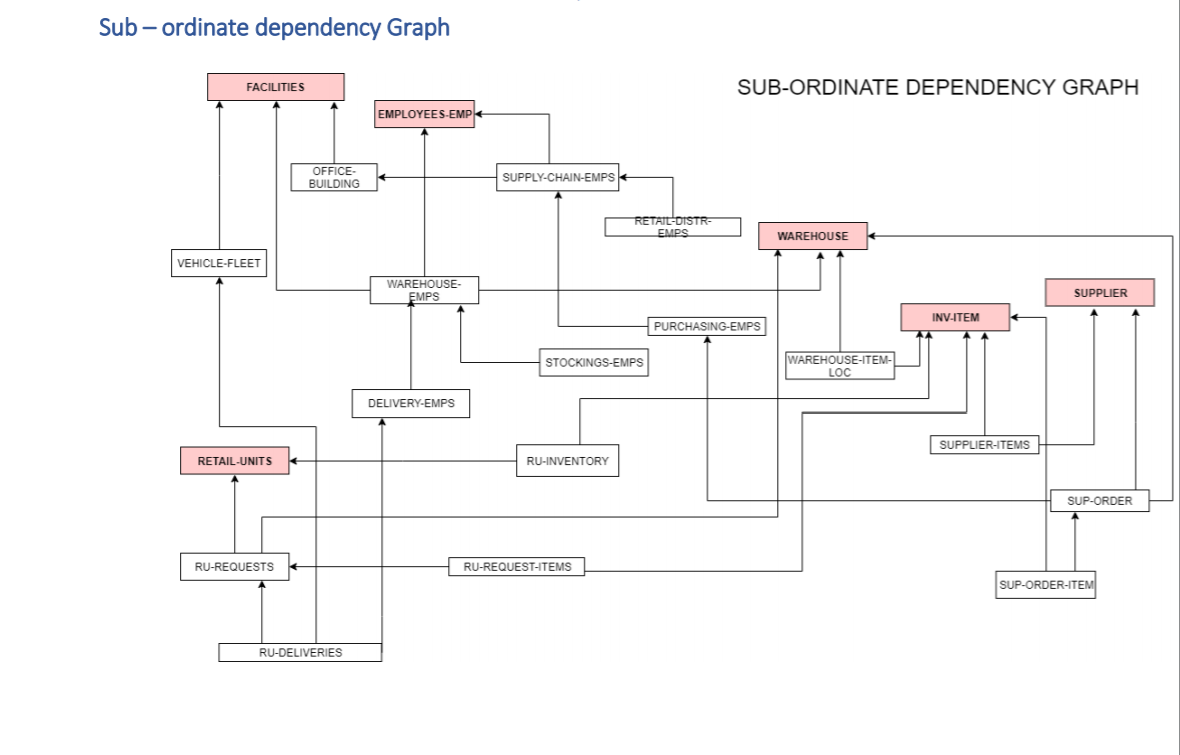
**SUPPLIER-STOCK.ITEM-ID 🡪 INV-ITEMS.ITEM-ID**

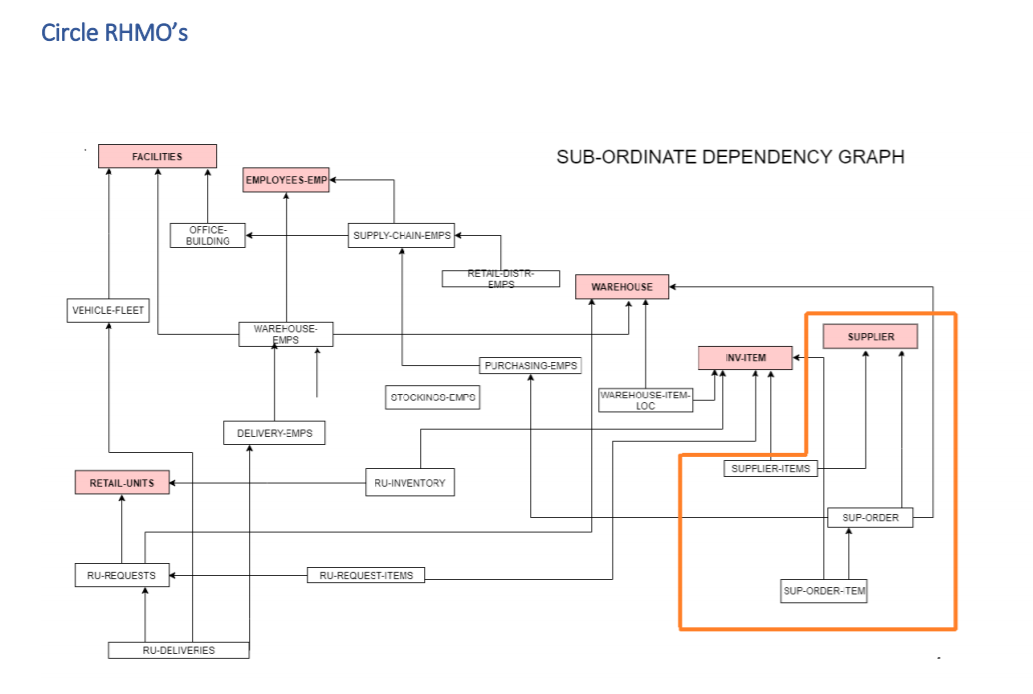
**CUST-ORDERS. S-EMP# 🡪 SALES-EMP.S-EMP#**

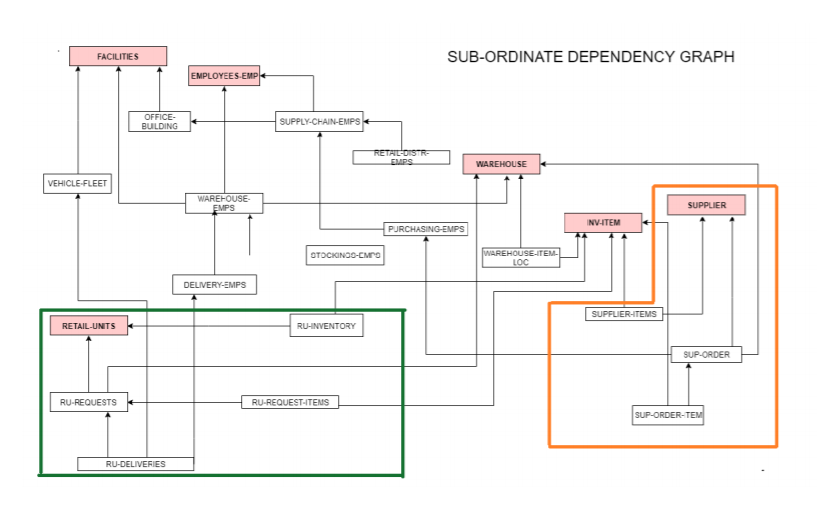
**CUST-ORDERS.CUST-ID 🡪 CUSTOMERS.CUST-ID**

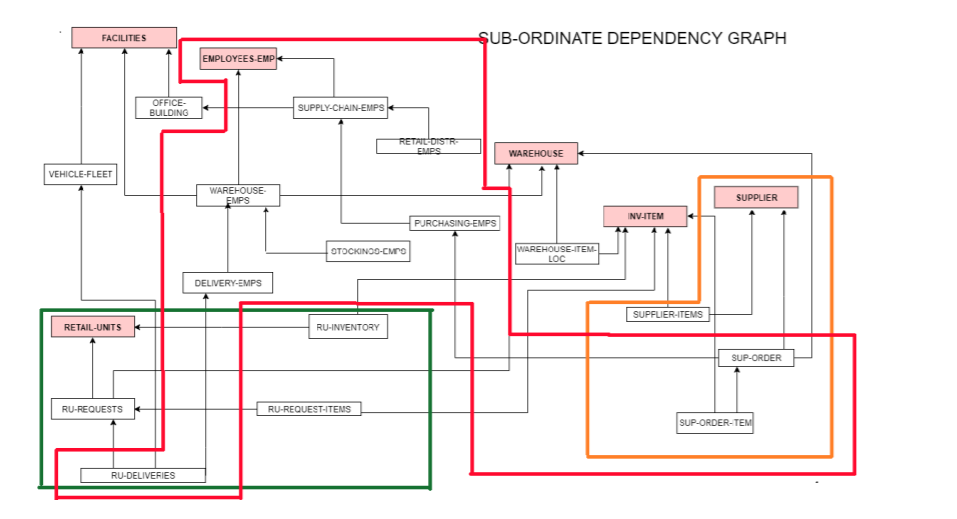
**CUST-ORDER-ITEMS.CUST-ORDER# 🡪 CUST-ORDERS.CUST-ORDER#**

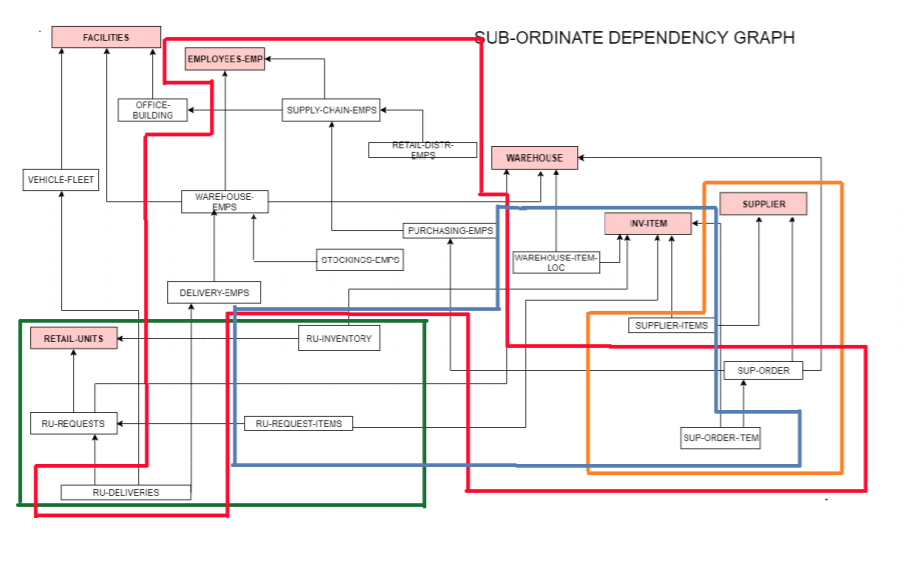
**CUST-ORDER-ITEMS.ITEM-ID 🡪 INV-ITEMS.ITEM-ID**

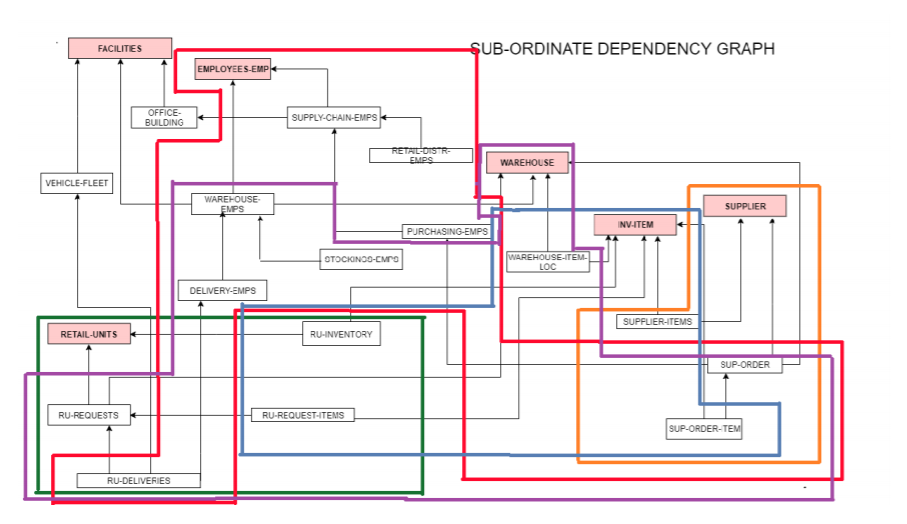


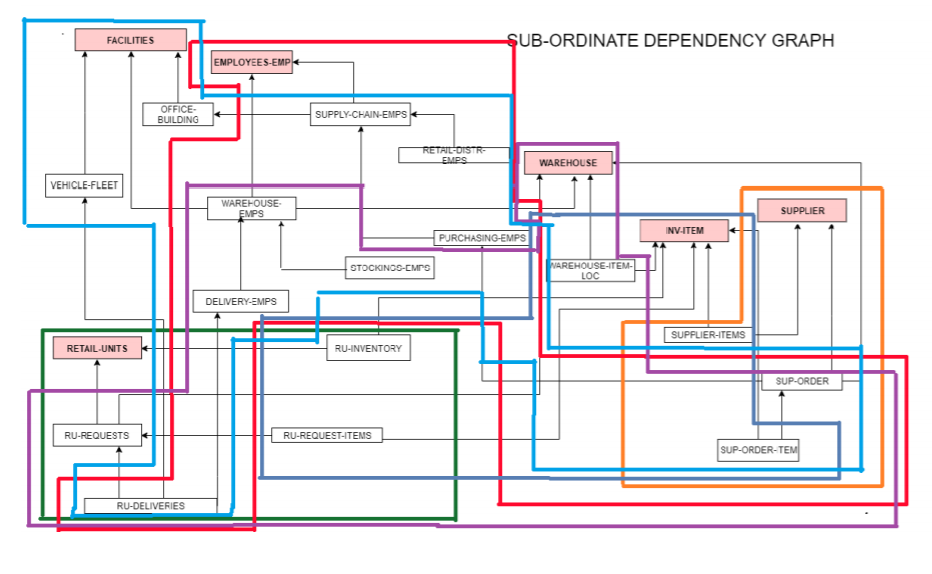


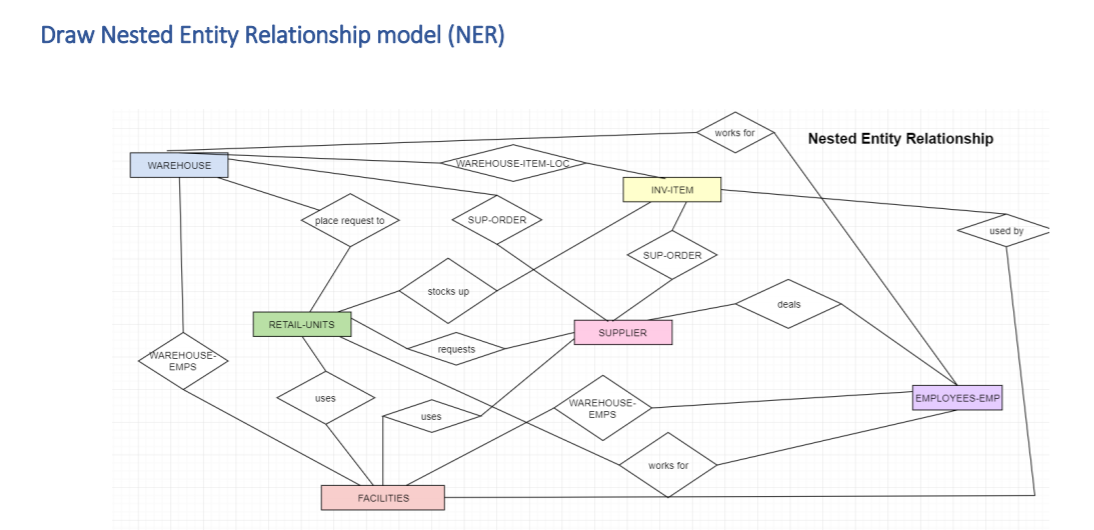


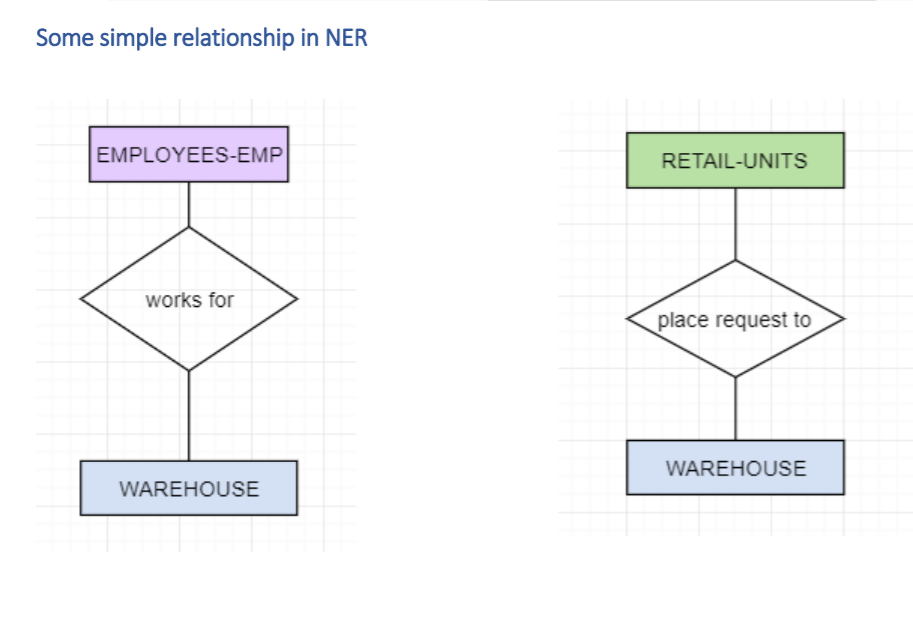


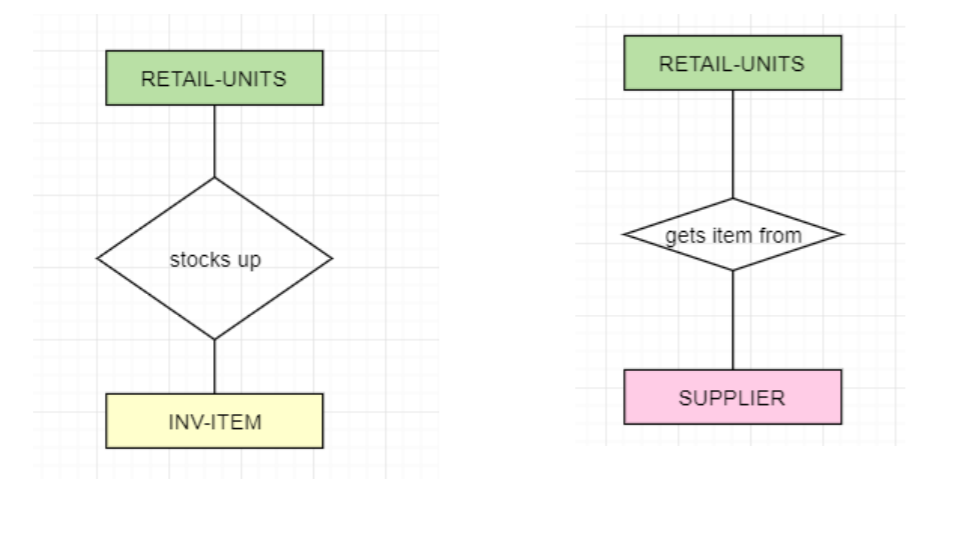


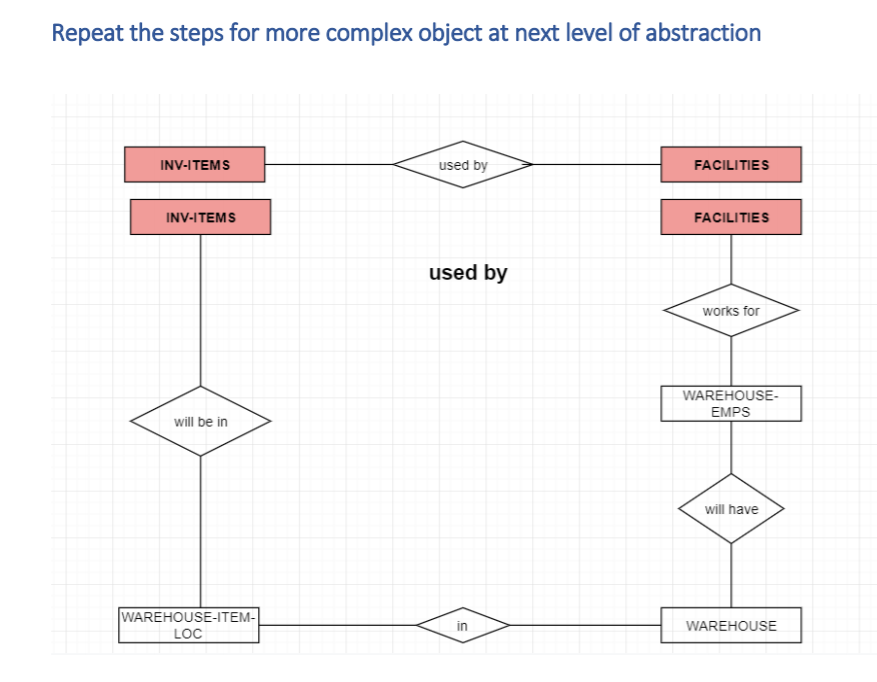


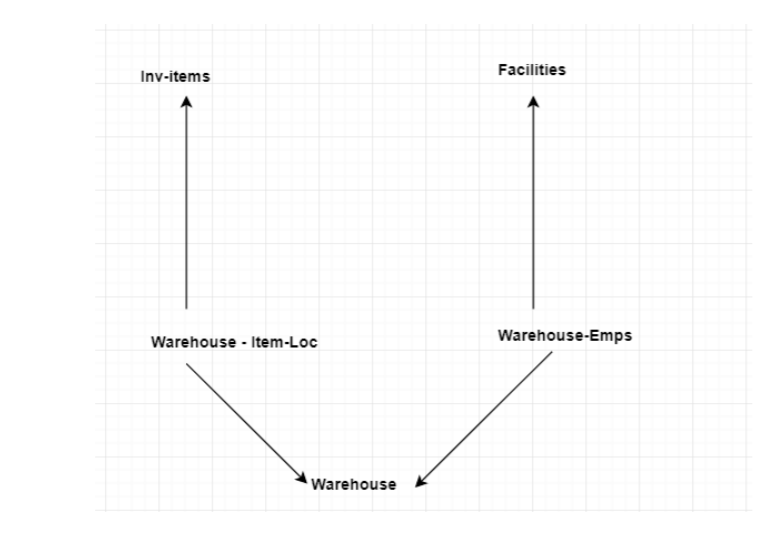


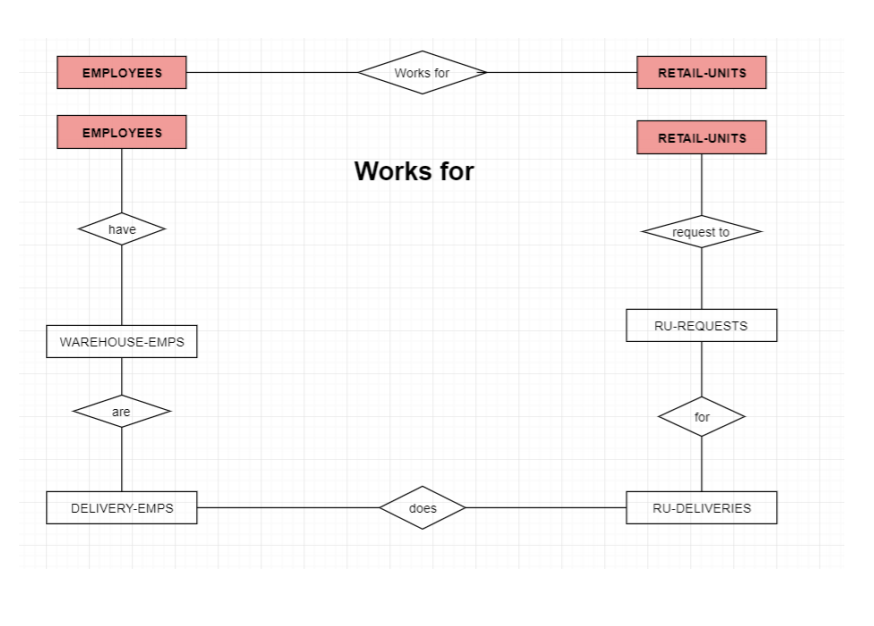


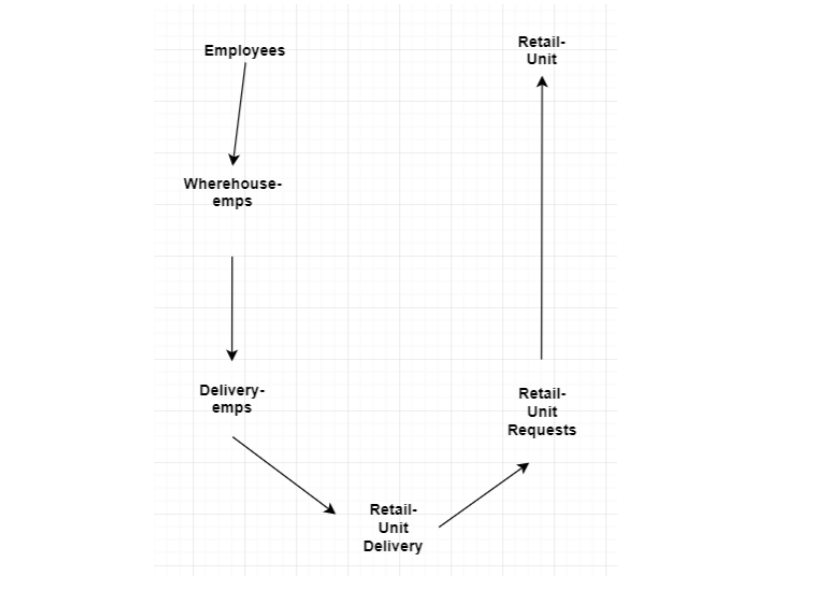


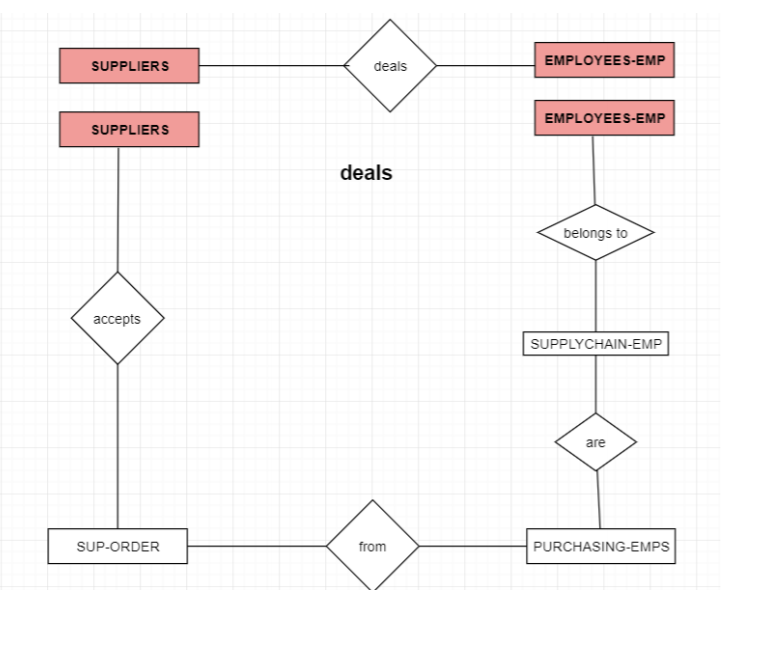


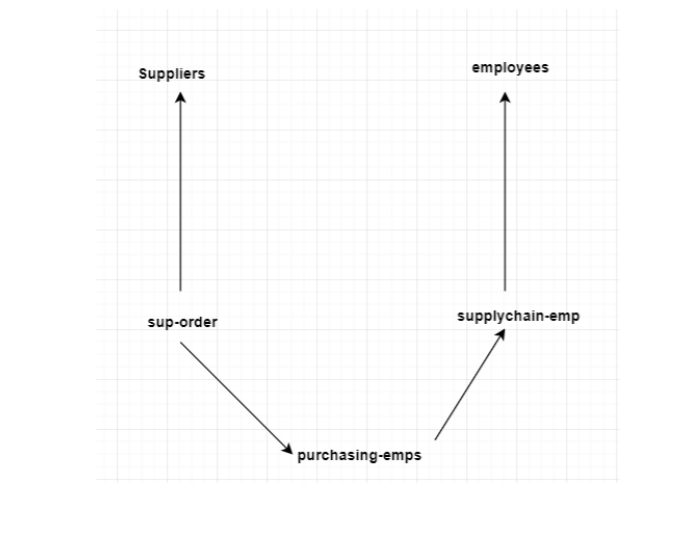


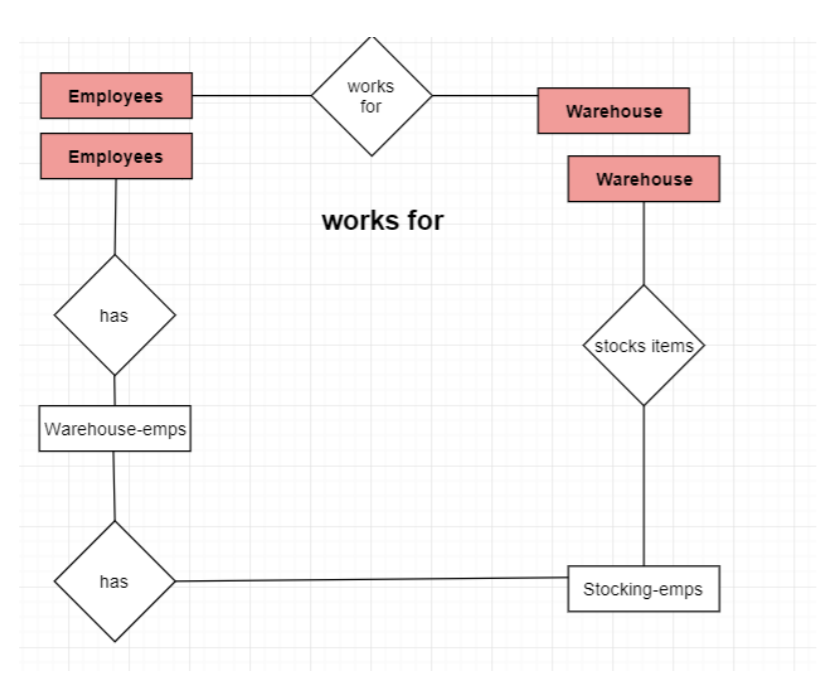


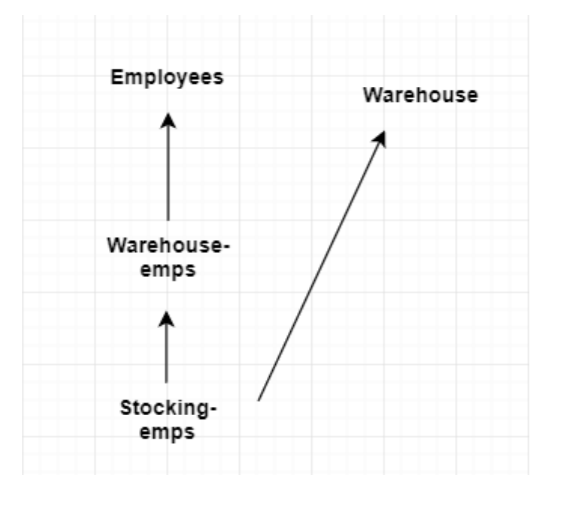


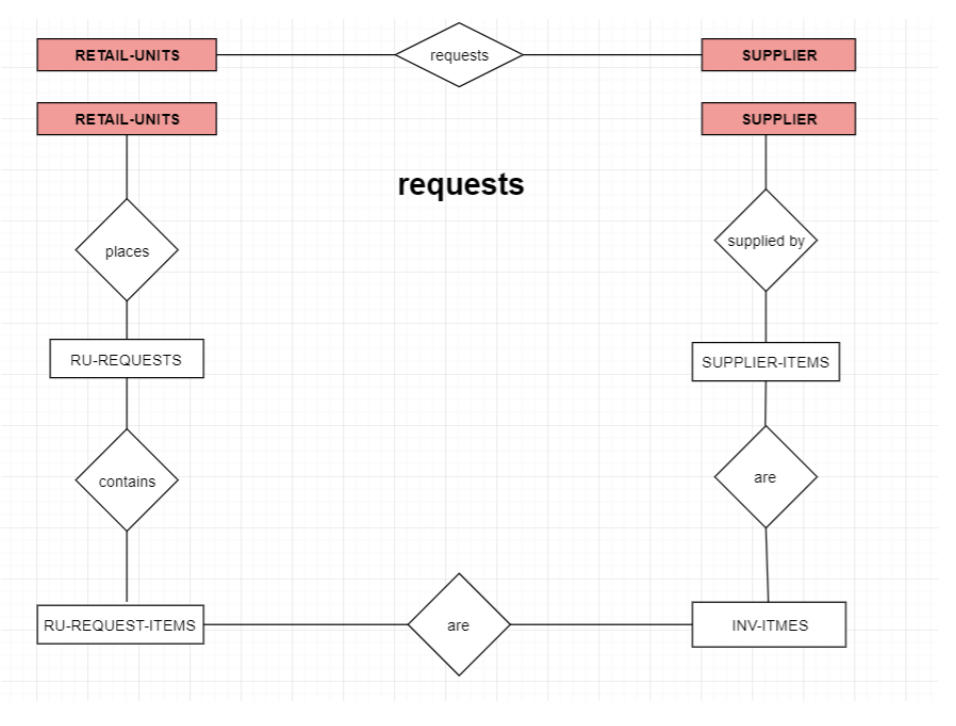


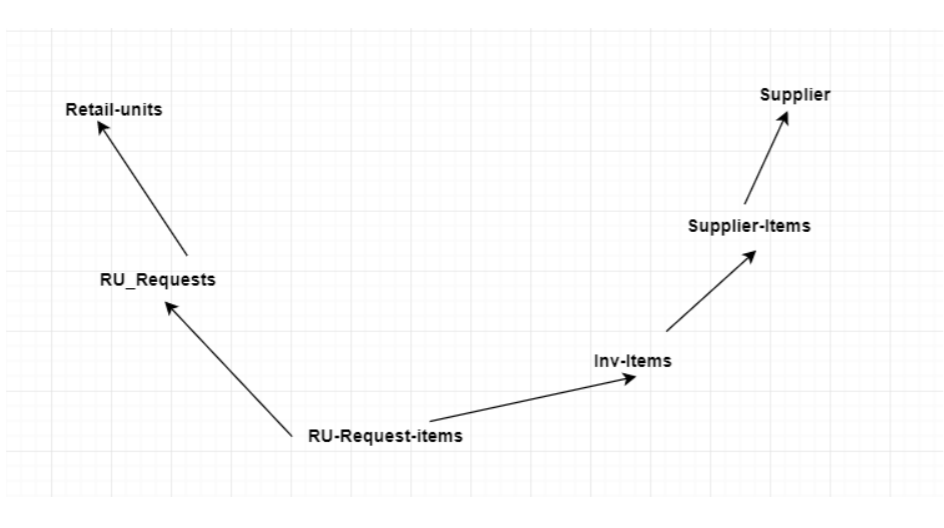












**2. Given the following Database and Attribute Value Classification Schema, generate the following four rules for COLRANK:**

**Characteristic Rules: Classification Rules:**

**TOP50 🡪 TOP50 🡨**

**BOT50 🡪 BOT50 🡨**

**Provide tables after the complete data substitution step and after the simplification step. Include count numbers for each profile.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **NAME** | **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| JANE | CLASS1 | TOP10 | 26 | ITM | A | TOP50 |
| JILL | CLASS2 | TOP50 | 18 | ART | C | BOT50 |
| JACK | CLASS1 | TOP25 | 29 | ME | B | TOP50 |
| KIM | CLASS2 | TOP10 | 25 | ITM | B | TOP50 |
| BILL | CLASS2 | TOP25 | 24 | EE | C | BOT50 |
| CRIS | CLASS2 | TOP25 | 24 | ITM | B | BOT50 |
| JIM | CLASS2 | TOP25 | 26 | ITM | A | TOP50 |
| LEE | CLASS1 | TOP25 | 24 | CE | C | TOP50 |
| DON | CLASS1 | TOP50 | 28 | EE | B | TOP50 |
| EARL | CLASS2 | TOP10 | 24 | BUS | C | BOT50 |
| MIRA | CLASS2 | TOP10 | 32 | SS | A | TOP50 |
| TONI | CLASS2 | TOP25 | 22 | CE | C | BOT50 |
| LEN | CLASS1 | TOP25 | 20 | BUS | B | BOT50 |
| ALLIE | CLASS1 | TOP10 | 33 | PREMED | A | TOP50 |
| CRISTEN | CLASS1 | TOP25 | 32 | PREMED | B | TOP50 |
| PETE | CLASS1 | TOP50 | 32 | HUM | B | TOP50 |
| JAY | CLASS1 | TOP50 | 25 | ME | B | TOP50 |
| KATE | CLASS2 | TOP50 | 21 | HUM | C | BOT50 |
| BRUCE | CLASS2 | TOP50 | 24 | SS | C | BOT50 |
| STU | CLASS1 | TOP50 | 28 | SS | B | BOT50 |

ATRIBUTE VALUE CLASSIFICATION SCHEMA

**CLSRANK:**

(TOP10, TOP25) A: ACCEPTABLE

(TOP50) Q: QUESTIONABLE

**ACT:**

(12-25) Q: QUESTIONABLE

(26-36) A: ACCEPTABLE

**MAJOR:**

(HUM, PREMED, SS, ART, MIS, BUS) SOL: SCHOOL OF LETTERS

(MATH, PHYS, ITM, EE, CE, ME) SOT: SCHOOL OF TECHNOLOGY

**ENTTEST:**

(A, B) A: ACCEPTABLE

(C, D) Q: QUESTIONABLE

Solution:

Characteristic Rules:

**TOP50:**

*Step1: Focus group extracted from the given table for ‘TOP50’ :*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **NAME** | **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| JANE | CLASS1 | TOP10 | 26 | ITM | A | TOP50 |
| JACK | CLASS1 | TOP25 | 29 | ME | B | TOP50 |
| KIM | CLASS2 | TOP10 | 25 | ITM | B | TOP50 |
| JIM | CLASS2 | TOP25 | 26 | ITM | A | TOP50 |
| LEE | CLASS1 | TOP25 | 24 | CE | C | TOP50 |
| DON | CLASS1 | TOP50 | 28 | EE | B | TOP50 |
| MIRA | CLASS2 | TOP10 | 32 | SS | A | TOP50 |
| ALLIE | CLASS1 | TOP10 | 33 | PREMED | A | TOP50 |
| CRISTEN | CLASS1 | TOP25 | 32 | PREMED | B | TOP50 |
| PETE | CLASS1 | TOP50 | 32 | HUM | B | TOP50 |
| JAY | CLASS1 | TOP50 | 25 | ME | B | TOP50 |

As seen in the above table, from the given total of 20 profiles, 11 profiles have COLRANK as TOP50.

*Step2: Generalized relation after data substitution from the given schema and removing NAME and COLRANK attributes*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** |
| CLASS1 | A | A | SOT | A |
| CLASS1 | A | A | SOT | A |
| CLASS2 | A | Q | SOT | A |
| CLASS2 | A | A | SOT | A |
| CLASS1 | A | Q | SOT | Q |
| CLASS1 | Q | A | SOT | A |
| CLASS2 | A | A | SOL | A |
| CLASS1 | A | A | SOL | A |
| CLASS1 | A | A | SOL | A |
| CLASS1 | Q | A | SOL | A |
| CLASS1 | Q | Q | SOT | A |

*Step3:Highlighting same profiles*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** |
| CLASS1 | A | A | SOT | A |
| CLASS1 | A | A | SOT | A |
| CLASS2 | A | Q | SOT | A |
| CLASS2 | A | A | SOT | A |
| CLASS1 | A | Q | SOT | Q |
| CLASS1 | Q | A | SOT | A |
| CLASS2 | A | A | SOL | A |
| CLASS1 | A | A | SOL | A |
| CLASS1 | A | A | SOL | A |
| CLASS1 | Q | A | SOL | A |
| CLASS1 | Q | Q | SOT | A |

*Step4: Combining same profiles and updating count for every profile to get further generalized relation*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** |
| CLASS1 | A | A | SOT | A | 2 |
| CLASS2 | A | Q | SOT | A | 1 |
| CLASS2 | A | A | SOT | A | 1 |
| CLASS1 | A | Q | SOT | Q | 1 |
| CLASS1 | Q | A | SOT | A | 1 |
| CLASS2 | A | A | SOL | A | 1 |
| CLASS1 | A | A | SOL | A | 2 |
| CLASS1 | Q | A | SOL | A | 1 |
| CLASS1 | Q | Q | SOT | A | 1 |

*Step5:Highlighting the same profiles*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** |
| CLASS1 | A | A | SOT | A | 2 |
| CLASS2 | A | Q | SOT | A | 1 |
| CLASS2 | A | A | SOT | A | 1 |
| CLASS1 | A | Q | SOT | Q | 1 |
| CLASS1 | Q | A | SOT | A | 1 |
| CLASS2 | A | A | SOL | A | 1 |
| CLASS1 | A | A | SOL | A | 2 |
| CLASS1 | Q | A | SOL | A | 1 |
| CLASS1 | Q | Q | SOT | A | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** |
| CLASS2 | A | ANY | SOT | A | 3 |
| CLASS1 | A | Q | SOT | Q | 1 |
| CLASS1 | Q | A | SOT | A | 1 |
| CLASS2 | A | A | SOL | A | 1 |
| CLASS1 | ANY | A | SOL | A | 3 |
| CLASS1 | Q | Q | SOT | A | 1 |
| HIGHSCH | CLSRANK | ACT | MAJOR | ENTTEST | COUNT |
| CLASS2 | A | ANY | SOT | A | 3 |
| CLASS1 | A | Q | SOT | Q | 1 |
| CLASS1 | Q | A | SOT | A | 1 |
| CLASS2 | A | A | SOL | A | 1 |
| CLASS1 | ANY | A | SOL | A | 3 |
| CLASS1 | Q | Q | SOT | A | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** |
| CLASS2 | A | ANY | SOT | A | 3 |
| CLASS1 | A | Q | SOT | Q | 1 |
| CLASS1 | Q | ANY | SOT | A | 2 |
| CLASS2 | A | A | SOL | A | 1 |
| CLASS1 | ANY | A | SOL | A | 3 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** | **Percentage** |
| CLASS2 | A | ANY | SOT | A | 3 | 30 |
| CLASS1 | A | Q | SOT | Q | 1 | 10 |
| CLASS1 | Q | ANY | SOT | A | 2 | 20 |
| CLASS2 | A | A | SOL | A | 1 | 10 |
| CLASS1 | ANY | A | SOL | A | 3 | 30 |

*Step6:*

*Learning Characteristic Rules:*

Ɐ (X) TOP50 (X)--> (

(HIGHSCH(X) € CLASS2 ^ CLSRANK (X) € A ^ MAJOR(X) € SOT^ ENTTEST(X) € A )

(HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € A ^ ACT(X) € Q ^ MAJOR(X) € SOT ^ ENTTEST(X) € Q) (HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € Q ^ MAJOR(X) € SOT ^ ENTTEST(X) € A)

(HIGHSCH(X) € CLASS2 ^ CLSRANK (X) € A ^ ACT(X) € A ^ MAJOR(X) € SOL ^ ENTTEST(X) € A) V (HIGHSCH(X) € CLASS1 ^ ACT(X) € A ^ MAJOR(X) € SOL ^ ENTTEST(X) € A) )

*BOT50:*

*Step1:*

*Focus group extracted from the given table for ‘BOT50’ :*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **NAME** | **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| JILL | CLASS2 | TOP50 | 18 | ART | C | BOT50 |
| BILL | CLASS2 | TOP25 | 24 | EE | C | BOT50 |
| CRIS | CLASS2 | TOP25 | 24 | ITM | B | BOT50 |
| EARL | CLASS2 | TOP10 | 24 | BUS | C | BOT50 |
| TONI | CLASS2 | TOP25 | 22 | CE | C | BOT50 |
| LEN | CLASS1 | TOP25 | 20 | BUS | B | BOT50 |
| KATE | CLASS2 | TOP50 | 21 | HUM | C | BOT50 |
| BRUCE | CLASS2 | TOP50 | 24 | SS | C | BOT50 |
| STU | CLASS1 | TOP50 | 28 | SS | B | BOT50 |

As seen in the above table, from the given total of 20 profiles, 10 profiles have COLRANK as BOT50

*Step2: Generalized relation after data substitution from the given schema and removing NAME and COLRANK attributes.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** |
| CLASS2 | TOP50 | 18 | ART | C |
| CLASS2 | TOP25 | 24 | EE | C |
| CLASS2 | TOP25 | 24 | ITM | B |
| CLASS2 | TOP10 | 24 | BUS | C |
| CLASS2 | TOP25 | 22 | CE | C |
| CLASS1 | TOP25 | 20 | BUS | B |
| CLASS2 | TOP50 | 21 | HUM | C |
| CLASS2 | TOP50 | 24 | SS | C |
| CLASS1 | TOP50 | 28 | SS | B |

*Step3: Highlighting same profiles :*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** |
| CLASS2 | Q | Q | SOL | Q |
| CLASS2 | A | Q | SOT | Q |
| CLASS2 | A | Q | SOT | A |
| CLASS2 | A | Q | SOL | Q |
| CLASS2 | A | Q | SOT | Q |
| CLASS1 | A | Q | SOL | A |
| CLASS2 | Q | Q | SOL | Q |
| CLASS2 | Q | Q | SOL | Q |
| CLASS1 | Q | A | SOL | A |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** |
| CLASS2 | Q | Q | SOL | Q |
| CLASS2 | A | Q | SOT | Q |
| CLASS2 | A | Q | SOT | A |
| CLASS2 | A | Q | SOL | Q |
| CLASS2 | A | Q | SOT | Q |
| CLASS1 | A | Q | SOL | A |
| CLASS2 | Q | Q | SOL | Q |
| CLASS2 | Q | Q | SOL | Q |
| CLASS1 | Q | A | SOL | A |

*Step4: Combining same profiles and updating count for every profile to get further generalized relation*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** |
| CLASS2 | Q | Q | SOL | Q | 1 |
| CLASS2 | A | Q | SOT | Q | 2 |
| CLASS2 | A | Q | SOT | A | 1 |
| CLASS2 | A | Q | SOL | Q | 1 |
| CLASS1 | A | Q | SOL | A | 1 |
| CLASS2 | Q | Q | SOL | Q | 2 |
| CLASS1 | Q | A | SOL | A | 1 |

*Step5: ANY substitution – Highlighting common profiles with one different attribute – to get the simplified generalized relation*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** |
| CLASS2 | Q | Q | SOL | Q | 1 |
| CLASS2 | A | Q | SOT | Q | 2 |
| CLASS2 | A | Q | SOT | A | 1 |
| CLASS2 | A | Q | SOL | Q | 1 |
| CLASS1 | A | Q | SOL | A | 1 |
| CLASS2 | Q | Q | SOL | Q | 2 |
| CLASS1 | Q | A | SOL | A | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** | **Percentage** |
| CLASS2 | Q | Q | SOL | Q | 3 | 33.30% |
| CLASS2 | A | Q | SOT | ANY | 3 | 33.30% |
| CLASS2 | A | Q | SOL | Q | 1 | 11.10% |
| CLASS1 | A | Q | SOL | A | 1 | 11.10% |
| CLASS1 | Q | A | SOL | A | 1 | 11.10% |

*Step6:*

*Learning Characteristic Rules:*

Ɐ (X) BOT50 (X) -->

(

(HIGHSCH(X) € CLASS2 ^ CLSRANK (X) € Q ^ ACT(X) € Q ^ MAJOR(X) € SOL ^ ENTTEST(X) € Q )

(HIGHSCH(X) € CLASS2 ^ CLSRANK (X) € A ^ ACT(X) € Q ^ MAJOR(X) € SOT)

(HIGHSCH(X) € CLASS2 ^ CLSRANK (X) € A ^ ACT(X) € Q ^ MAJOR(X) € SOL ^ ENTTEST(X) € Q) (HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € A ^ ACT(X) € Q ^ MAJOR(X) € SOL ^ ENTTEST(X) € A) (HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € Q ^ ACT(X) € A ^ MAJOR(X) € SOL ^ ENTTEST(X) € A)

)

*Learning Classification Rules:*

*Step1: Generalized relation for both TOP50 and BOT50*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Learning Concept** | **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** |
| TOP50 | CLASS1 | A | A | SOT | A | 2 |
| CLASS2 | A | Q | SOT | A | 1 |
| CLASS2 | A | A | SOT | A | 1 |
| CLASS1 | A | Q | SOT | Q | 1 |
| CLASS1 | Q | A | SOT | A | 1 |
| CLASS2 | A | A | SOL | A | 1 |
| CLASS1 | A | A | SOL | A | 2 |
| CLASS1 | Q | A | SOL | A | 1 |
| CLASS1 | Q | Q | SOT | A | 1 |
| BOT50 | CLASS2 | Q | Q | SOL | Q | 3 |
| CLASS2 | A | Q | SOT | Q | 2 |
| CLASS2 | A | Q | SOT | A | 1 |
| CLASS2 | A | Q | SOL | Q | 1 |
| CLASS1 | A | Q | SOL | A | 1 |
| CLASS1 | Q | A | SOL | A | 1 |

*Step2: Identifying common profiles between both ranks*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Learning Concept** | **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** | **Mark** |
| TOP50 | CLASS1 | A | A | SOT | A | 2 |  |
| CLASS2 | A | Q | SOT | A | 1 | \* |
| CLASS2 | A | A | SOT | A | 1 |  |
| CLASS1 | A | Q | SOT | Q | 1 |  |
| CLASS1 | Q | A | SOT | A | 1 |  |
| CLASS2 | A | A | SOL | A | 1 |  |
| CLASS1 | A | A | SOL | A | 2 |  |
| CLASS1 | Q | A | SOL | A | 1 | \*\* |
| CLASS1 | Q | Q | SOT | A | 1 |  |
| BOT50 | CLASS2 | Q | Q | SOL | Q | 3 |  |
| CLASS2 | A | Q | SOT | Q | 2 |  |
| CLASS2 | A | Q | SOT | A | 1 | \* |
| CLASS2 | A | Q | SOL | Q | 1 |  |
| CLASS1 | A | Q | SOL | A | 1 |  |
| CLASS1 | Q | A | SOL | A | 1 | \*\* |

Step3: ANY substitution for common profiles with one different attribute and striking off the common profile between TOP50 and BOT50.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Learning Concept** | **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COUNT** |
| TOP50 | ANY | A | A | SOT | A | 3 |
| ~~CLASS2~~ | ~~A~~ | ~~Q~~ | ~~SOT~~ | ~~A~~ | ~~1~~ |
| CLASS1 | A | Q | SOT | Q | 1 |
| CLASS1 | Q | A | ANY | A | 2 |
| ANY | A | A | SOL | A | 3 |
| CLASS1 | Q | Q | SOT | A | 1 |
| BOT50 | CLASS2 | ANY | Q | SOL | Q | 4 |
| CLASS2 | A | Q | SOT | Q | 2 |
| ~~CLASS2~~ | ~~A~~ | ~~Q~~ | ~~SOT~~ | ~~A~~ | ~~1~~ |
| CLASS1 | A | Q | SOL | A | 1 |
| CLASS1 | Q | A | SOL | A | 1 |

*Step4:*

*Learning Classification rules Rule for TOP50:*

Ɐ (X) TOP50 (X) <-- (

(CLSRANK (X) € A ^ ACT(X) € A ^ MAJOR(X) € SOT ^ ENTTEST(X) € A)

(HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € A ^ ACT(X) € Q ^ MAJOR(X) € SOT ^ ENTTEST(X) € Q)

(HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € Q ^ ACT(X) € A ^ ENTTEST(X) € A)

(CLSRANK (X) € A ^ ACT(X) € A ^ MAJOR(X) € SOL ^ ENTTEST(X) € A)

(HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € Q ^ ACT(X) € Q ^ MAJOR(X) € SOT ^ ENTTEST(X) € A)

)

*Rule for BOT50:*

Ɐ (X) BOT50 (X) <-- (

(HIGHSCH(X) € CLASS2 ^ ACT(X) € Q ^ MAJOR(X) € SOL ^ ENTTEST(X) € Q)

(HIGHSCH(X) € CLASS2 ^ CLSRANK (X) € A ^ ACT(X) € Q^ MAJOR(X) € SOT ^ ENTTEST(X) € Q)

(HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € A ^ ACT(X) € Q ^ MAJOR(X) € SOL ^ ENTTEST(X) € A)

(HIGHSCH(X) € CLASS1 ^ CLSRANK (X) € Q ^ ACT(X) € A ^ MAJOR(X) € SOL ^ ENTTEST(X) € A)

)

**3. Using the table in problem #1,**

* 1. **Develop a decision tree based on the method discussed in class which uses HIGHSCH, and CLSRANK information as a possible predictor of who will end up in the TOP50 or BOT50 in their COLRANK.**

*Generalized Relations:*

*HIGHSCH: CLASS1*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | TOP10 | 26 | ITM | A | TOP50 |
| CLASS2 | TOP50 | 18 | ART | C | BOT50 |
| CLASS1 | TOP25 | 29 | ME | B | TOP50 |
| CLASS2 | TOP10 | 25 | ITM | B | TOP50 |
| CLASS2 | TOP25 | 24 | EE | C | BOT50 |
| CLASS2 | TOP25 | 24 | ITM | B | BOT50 |
| CLASS2 | TOP25 | 26 | ITM | A | TOP50 |
| CLASS1 | TOP25 | 24 | CE | C | TOP50 |
| CLASS1 | TOP50 | 28 | EE | B | TOP50 |
| CLASS2 | TOP10 | 24 | BUS | C | BOT50 |
| CLASS2 | TOP10 | 32 | SS | A | TOP50 |
| CLASS2 | TOP25 | 22 | CE | C | BOT50 |
| CLASS1 | TOP25 | 20 | BUS | B | BOT50 |
| CLASS1 | TOP10 | 33 | PREMED | A | TOP50 |
| CLASS1 | TOP25 | 32 | PREMED | B | TOP50 |
| CLASS1 | TOP50 | 32 | HUM | B | TOP50 |
| CLASS1 | TOP50 | 25 | ME | B | TOP50 |
| CLASS2 | TOP50 | 21 | HUM | C | BOT50 |
| CLASS2 | TOP50 | 24 | SS | C | BOT50 |
| CLASS1 | TOP50 | 28 | SS | B | BOT50 |

*CLASS2:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | TOP10 | 26 | ITM | A | TOP50 |
| CLASS2 | TOP50 | 18 | ART | C | BOT50 |
| CLASS1 | TOP25 | 29 | ME | B | TOP50 |
| CLASS2 | TOP10 | 25 | ITM | B | TOP50 |
| CLASS2 | TOP25 | 24 | EE | C | BOT50 |
| CLASS2 | TOP25 | 24 | ITM | B | BOT50 |
| CLASS2 | TOP25 | 26 | ITM | A | TOP50 |
| CLASS1 | TOP25 | 24 | CE | C | TOP50 |
| CLASS1 | TOP50 | 28 | EE | B | TOP50 |
| CLASS2 | TOP10 | 24 | BUS | C | BOT50 |
| CLASS2 | TOP10 | 32 | SS | A | TOP50 |
| CLASS2 | TOP25 | 22 | CE | C | BOT50 |
| CLASS1 | TOP25 | 20 | BUS | B | BOT50 |
| CLASS1 | TOP10 | 33 | PREMED | A | TOP50 |
| CLASS1 | TOP25 | 32 | PREMED | B | TOP50 |
| CLASS1 | TOP50 | 32 | HUM | B | TOP50 |
| CLASS1 | TOP50 | 25 | ME | B | TOP50 |
| CLASS2 | TOP50 | 21 | HUM | C | BOT50 |
| CLASS2 | TOP50 | 24 | SS | C | BOT50 |
| CLASS1 | TOP50 | 28 | SS | B | BOT50 |

*Predictability :*

|  |  |  |
| --- | --- | --- |
| HIGHSCH | | |
| COLRANK | CLASS1 | CLASS2 |
| TOP50 | 8 | 3 |
| BOT50 | 2 | 7 |

Predictability of using CLASS1 implies TOP50 and CLASS2 implies BOT50.

We would be

RIGHT : (8 + 7 ) / ( 8 + 7 + 2 + 3) : 15/20 times

WRONG : (2+ 3 ) / ( 8 + 7 + 2 + 3) : 5/20 times

*CLSRANK: Generalized Relation:*

*CLSRANK : A*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | 26 | ITM | A | TOP50 |
| CLASS2 | Q | 18 | ART | C | BOT50 |
| CLASS1 | A | 29 | ME | B | TOP50 |
| CLASS2 | A | 25 | ITM | B | TOP50 |
| CLASS2 | A | 24 | EE | C | BOT50 |
| CLASS2 | A | 24 | ITM | B | BOT50 |
| CLASS2 | A | 26 | ITM | A | TOP50 |
| CLASS1 | A | 24 | CE | C | TOP50 |
| CLASS1 | Q | 28 | EE | B | TOP50 |
| CLASS2 | A | 24 | BUS | C | BOT50 |
| CLASS2 | A | 32 | SS | A | TOP50 |
| CLASS2 | A | 22 | CE | C | BOT50 |
| CLASS1 | A | 20 | BUS | B | BOT50 |
| CLASS1 | A | 33 | PREMED | A | TOP50 |
| CLASS1 | A | 32 | PREMED | B | TOP50 |
| CLASS1 | Q | 32 | HUM | B | TOP50 |
| CLASS1 | Q | 25 | ME | B | TOP50 |
| CLASS2 | Q | 21 | HUM | C | BOT50 |
| CLASS2 | Q | 24 | SS | C | BOT50 |
| CLASS1 | Q | 28 | SS | B | BOT50 |

*CLSRANK : Q*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | 26 | ITM | A | TOP50 |
| CLASS2 | Q | 18 | ART | C | BOT50 |
| CLASS1 | A | 29 | ME | B | TOP50 |
| CLASS2 | A | 25 | ITM | B | TOP50 |
| CLASS2 | A | 24 | EE | C | BOT50 |
| CLASS2 | A | 24 | ITM | B | BOT50 |
| CLASS2 | A | 26 | ITM | A | TOP50 |
| CLASS1 | A | 24 | CE | C | TOP50 |
| CLASS1 | Q | 28 | EE | B | TOP50 |
| CLASS2 | A | 24 | BUS | C | BOT50 |
| CLASS2 | A | 32 | SS | A | TOP50 |
| CLASS2 | A | 22 | CE | C | BOT50 |
| CLASS1 | A | 20 | BUS | B | BOT50 |
| CLASS1 | A | 33 | PREMED | A | TOP50 |
| CLASS1 | A | 32 | PREMED | B | TOP50 |
| CLASS1 | Q | 32 | HUM | B | TOP50 |
| CLASS1 | Q | 25 | ME | B | TOP50 |
| CLASS2 | Q | 21 | HUM | C | BOT50 |
| CLASS2 | Q | 24 | SS | C | BOT50 |
| CLASS1 | Q | 28 | SS | B | BOT50 |

Predictablility:

|  |  |  |
| --- | --- | --- |
| CLSRANK | | |
| COLRANK | A | Q |
| TOP50 | 8 | 3 |
| BOT50 | 5 | 4 |

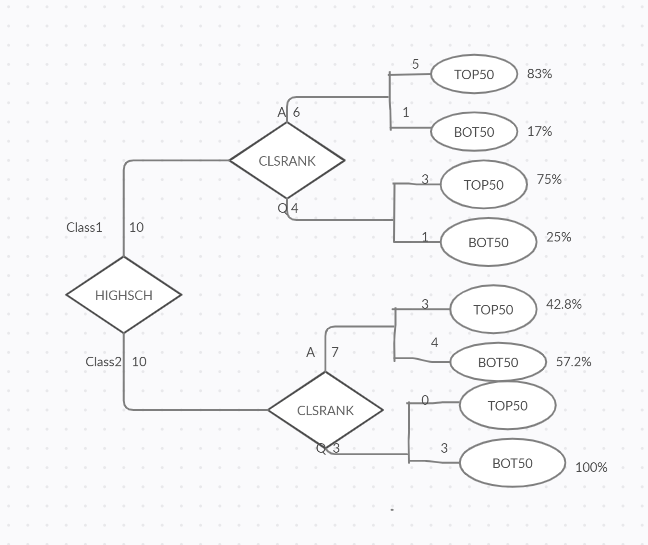
Predictability of using A implies TOP50 and Q implies BOT50

We would be RIGHT: (8 + 4 ) / ( 8 + 4 + 3 + 5) : 12/20 times

WRONG: (5 + 3 ) / ( 8 + 4 + 3 + 5) : 8/20 times

The attribute ‘HIGHSCH’ is taken as the root node considering the high predictability value.

*Decision Tree:*



**b.Develop a second decision tree using HIGHSCH, CLSRANK, ACT, MAJOR, and ENTTEST information as a possible predictor of who will end up in the TOP50 and BOT50 in their COLRANK.**

*Generalized Relations:*

*HIGHSCH: CLASS1*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | TOP10 | 26 | ITM | A | TOP50 |
| CLASS2 | TOP50 | 18 | ART | C | BOT50 |
| CLASS1 | TOP25 | 29 | ME | B | TOP50 |
| CLASS2 | TOP10 | 25 | ITM | B | TOP50 |
| CLASS2 | TOP25 | 24 | EE | C | BOT50 |
| CLASS2 | TOP25 | 24 | ITM | B | BOT50 |
| CLASS2 | TOP25 | 26 | ITM | A | TOP50 |
| CLASS1 | TOP25 | 24 | CE | C | TOP50 |
| CLASS1 | TOP50 | 28 | EE | B | TOP50 |
| CLASS2 | TOP10 | 24 | BUS | C | BOT50 |
| CLASS2 | TOP10 | 32 | SS | A | TOP50 |
| CLASS2 | TOP25 | 22 | CE | C | BOT50 |
| CLASS1 | TOP25 | 20 | BUS | B | BOT50 |
| CLASS1 | TOP10 | 33 | PREMED | A | TOP50 |
| CLASS1 | TOP25 | 32 | PREMED | B | TOP50 |
| CLASS1 | TOP50 | 32 | HUM | B | TOP50 |
| CLASS1 | TOP50 | 25 | ME | B | TOP50 |
| CLASS2 | TOP50 | 21 | HUM | C | BOT50 |
| CLASS2 | TOP50 | 24 | SS | C | BOT50 |
| CLASS1 | TOP50 | 28 | SS | B | BOT50 |

*CLASS2:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | TOP10 | 26 | ITM | A | TOP50 |
| CLASS2 | TOP50 | 18 | ART | C | BOT50 |
| CLASS1 | TOP25 | 29 | ME | B | TOP50 |
| CLASS2 | TOP10 | 25 | ITM | B | TOP50 |
| CLASS2 | TOP25 | 24 | EE | C | BOT50 |
| CLASS2 | TOP25 | 24 | ITM | B | BOT50 |
| CLASS2 | TOP25 | 26 | ITM | A | TOP50 |
| CLASS1 | TOP25 | 24 | CE | C | TOP50 |
| CLASS1 | TOP50 | 28 | EE | B | TOP50 |
| CLASS2 | TOP10 | 24 | BUS | C | BOT50 |
| CLASS2 | TOP10 | 32 | SS | A | TOP50 |
| CLASS2 | TOP25 | 22 | CE | C | BOT50 |
| CLASS1 | TOP25 | 20 | BUS | B | BOT50 |
| CLASS1 | TOP10 | 33 | PREMED | A | TOP50 |
| CLASS1 | TOP25 | 32 | PREMED | B | TOP50 |
| CLASS1 | TOP50 | 32 | HUM | B | TOP50 |
| CLASS1 | TOP50 | 25 | ME | B | TOP50 |
| CLASS2 | TOP50 | 21 | HUM | C | BOT50 |
| CLASS2 | TOP50 | 24 | SS | C | BOT50 |
| CLASS1 | TOP50 | 28 | SS | B | BOT50 |

*Predictability :*

|  |  |  |
| --- | --- | --- |
| HIGHSCH | | |
| COLRANK | CLASS1 | CLASS2 |
| TOP50 | 8 | 3 |
| BOT50 | 2 | 7 |

Predictability of using CLASS1 implies TOP50 and CLASS2 implies BOT50.

We would be

RIGHT : (8 + 7 ) / ( 8 + 7 + 2 + 3) : 15/20 times

WRONG : (2+ 3 ) / ( 8 + 7 + 2 + 3) : 5/20 times

*CLSRANK: Generalized Relation:*

*CLSRANK : A*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | 26 | ITM | A | TOP50 |
| CLASS2 | Q | 18 | ART | C | BOT50 |
| CLASS1 | A | 29 | ME | B | TOP50 |
| CLASS2 | A | 25 | ITM | B | TOP50 |
| CLASS2 | A | 24 | EE | C | BOT50 |
| CLASS2 | A | 24 | ITM | B | BOT50 |
| CLASS2 | A | 26 | ITM | A | TOP50 |
| CLASS1 | A | 24 | CE | C | TOP50 |
| CLASS1 | Q | 28 | EE | B | TOP50 |
| CLASS2 | A | 24 | BUS | C | BOT50 |
| CLASS2 | A | 32 | SS | A | TOP50 |
| CLASS2 | A | 22 | CE | C | BOT50 |
| CLASS1 | A | 20 | BUS | B | BOT50 |
| CLASS1 | A | 33 | PREMED | A | TOP50 |
| CLASS1 | A | 32 | PREMED | B | TOP50 |
| CLASS1 | Q | 32 | HUM | B | TOP50 |
| CLASS1 | Q | 25 | ME | B | TOP50 |
| CLASS2 | Q | 21 | HUM | C | BOT50 |
| CLASS2 | Q | 24 | SS | C | BOT50 |
| CLASS1 | Q | 28 | SS | B | BOT50 |

*CLSRANK : Q*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | 26 | ITM | A | TOP50 |
| CLASS2 | Q | 18 | ART | C | BOT50 |
| CLASS1 | A | 29 | ME | B | TOP50 |
| CLASS2 | A | 25 | ITM | B | TOP50 |
| CLASS2 | A | 24 | EE | C | BOT50 |
| CLASS2 | A | 24 | ITM | B | BOT50 |
| CLASS2 | A | 26 | ITM | A | TOP50 |
| CLASS1 | A | 24 | CE | C | TOP50 |
| CLASS1 | Q | 28 | EE | B | TOP50 |
| CLASS2 | A | 24 | BUS | C | BOT50 |
| CLASS2 | A | 32 | SS | A | TOP50 |
| CLASS2 | A | 22 | CE | C | BOT50 |
| CLASS1 | A | 20 | BUS | B | BOT50 |
| CLASS1 | A | 33 | PREMED | A | TOP50 |
| CLASS1 | A | 32 | PREMED | B | TOP50 |
| CLASS1 | Q | 32 | HUM | B | TOP50 |
| CLASS1 | Q | 25 | ME | B | TOP50 |
| CLASS2 | Q | 21 | HUM | C | BOT50 |
| CLASS2 | Q | 24 | SS | C | BOT50 |
| CLASS1 | Q | 28 | SS | B | BOT50 |

*Predictability :*

|  |  |  |
| --- | --- | --- |
| CLSRANK | | |
| COLRANK | A | Q |
| TOP50 | 8 | 3 |
| BOT50 | 5 | 4 |

Predictability of using A implies TOP50 and Q implies BOT50

We would be RIGHT: (8 + 4 ) / ( 8 + 4 + 3 + 5) : 12/20 times

WRONG: (5 + 3 ) / ( 8 + 4 + 3 + 5) : 8/20 times

*ACT: Generalized Relation:*

*ACT: A*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | A | ITM | A | TOP50 |
| CLASS2 | Q | Q | ART | C | BOT50 |
| CLASS1 | A | A | ME | B | TOP50 |
| CLASS2 | A | Q | ITM | B | TOP50 |
| CLASS2 | A | Q | EE | C | BOT50 |
| CLASS2 | A | Q | ITM | B | BOT50 |
| CLASS2 | A | A | ITM | A | TOP50 |
| CLASS1 | A | Q | CE | C | TOP50 |
| CLASS1 | Q | A | EE | B | TOP50 |
| CLASS2 | A | Q | BUS | C | BOT50 |
| CLASS2 | A | A | SS | A | TOP50 |
| CLASS2 | A | Q | CE | C | BOT50 |
| CLASS1 | A | Q | BUS | B | BOT50 |
| CLASS1 | A | A | PREMED | A | TOP50 |
| CLASS1 | A | A | PREMED | B | TOP50 |
| CLASS1 | Q | A | HUM | B | TOP50 |
| CLASS1 | Q | Q | ME | B | TOP50 |
| CLASS2 | Q | Q | HUM | C | BOT50 |
| CLASS2 | Q | Q | SS | C | BOT50 |
| CLASS1 | Q | A | SS | B | BOT50 |

*ACT:Q*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | A | ITM | A | TOP50 |
| CLASS2 | Q | Q | ART | C | BOT50 |
| CLASS1 | A | A | ME | B | TOP50 |
| CLASS2 | A | Q | ITM | B | TOP50 |
| CLASS2 | A | Q | EE | C | BOT50 |
| CLASS2 | A | Q | ITM | B | BOT50 |
| CLASS2 | A | A | ITM | A | TOP50 |
| CLASS1 | A | Q | CE | C | TOP50 |
| CLASS1 | Q | A | EE | B | TOP50 |
| CLASS2 | A | Q | BUS | C | BOT50 |
| CLASS2 | A | A | SS | A | TOP50 |
| CLASS2 | A | Q | CE | C | BOT50 |
| CLASS1 | A | Q | BUS | B | BOT50 |
| CLASS1 | A | A | PREMED | A | TOP50 |
| CLASS1 | A | A | PREMED | B | TOP50 |
| CLASS1 | Q | A | HUM | B | TOP50 |
| CLASS1 | Q | Q | ME | B | TOP50 |
| CLASS2 | Q | Q | HUM | C | BOT50 |
| CLASS2 | Q | Q | SS | C | BOT50 |
| CLASS1 | Q | A | SS | B | BOT50 |

*Predictability :*

|  |  |  |
| --- | --- | --- |
| ACT | | |
| COLRANK | A | Q |
| TOP50 | 8 | 3 |
| BOT50 | 1 | 8 |

Predictability of using A implies TOP50 and Q implies BOT50

We would be

RIGHT: (8 + 8 ) / ( 8+1+3+8) : 16/20 times

WRONG:( 1 + 3 ) / ( 8+1+3+8: 4/20 times

*MAJOR:SOT*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | A | SOT | A | TOP50 |
| CLASS2 | Q | Q | SOL | C | BOT50 |
| CLASS1 | A | A | SOT | B | TOP50 |
| CLASS2 | A | Q | SOT | B | TOP50 |
| CLASS2 | A | Q | SOT | C | BOT50 |
| CLASS2 | A | Q | SOT | B | BOT50 |
| CLASS2 | A | A | SOT | A | TOP50 |
| CLASS1 | A | Q | SOT | C | TOP50 |
| CLASS1 | Q | A | SOT | B | TOP50 |
| CLASS2 | A | Q | SOL | C | BOT50 |
| CLASS2 | A | A | SOL | A | TOP50 |
| CLASS2 | A | Q | SOT | C | BOT50 |
| CLASS1 | A | Q | SOL | B | BOT50 |
| CLASS1 | A | A | SOL | A | TOP50 |
| CLASS1 | A | A | SOL | B | TOP50 |
| CLASS1 | Q | A | SOL | B | TOP50 |
| CLASS1 | Q | Q | SOT | B | TOP50 |
| CLASS2 | Q | Q | SOL | C | BOT50 |
| CLASS2 | Q | Q | SOL | C | BOT50 |
| CLASS1 | Q | A | SOL | B | BOT50 |

*MAJOR:SOL*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | A | SOT | A | TOP50 |
| CLASS2 | Q | Q | SOL | C | BOT50 |
| CLASS1 | A | A | SOT | B | TOP50 |
| CLASS2 | A | Q | SOT | B | TOP50 |
| CLASS2 | A | Q | SOT | C | BOT50 |
| CLASS2 | A | Q | SOT | B | BOT50 |
| CLASS2 | A | A | SOT | A | TOP50 |
| CLASS1 | A | Q | SOT | C | TOP50 |
| CLASS1 | Q | A | SOT | B | TOP50 |
| CLASS2 | A | Q | SOL | C | BOT50 |
| CLASS2 | A | A | SOL | A | TOP50 |
| CLASS2 | A | Q | SOT | C | BOT50 |
| CLASS1 | A | Q | SOL | B | BOT50 |
| CLASS1 | A | A | SOL | A | TOP50 |
| CLASS1 | A | A | SOL | B | TOP50 |
| CLASS1 | Q | A | SOL | B | TOP50 |
| CLASS1 | Q | Q | SOT | B | TOP50 |
| CLASS2 | Q | Q | SOL | C | BOT50 |
| CLASS2 | Q | Q | SOL | C | BOT50 |
| CLASS1 | Q | A | SOL | B | BOT50 |

Predictability :

|  |  |  |
| --- | --- | --- |
| MAJOR | | |
| COLRANK | SOT | SOL |
| TOP50 | 7 | 4 |
| BOT50 | 3 | 6 |

Predictability of using SOT implies TOP50 and SOL implies BOT50

We would be

RIGHT : (7 + 6 ) / ( 7 + 6 + 3 + 4) : 13/20 times

WRONG : (3 + 4 ) / ( 7 + 6 + 3 + 4) : 7/20 times

*ENTTEST-A*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | A | SOT | A | TOP50 |
| CLASS2 | Q | Q | SOL | Q | BOT50 |
| CLASS1 | A | A | SOT | A | TOP50 |
| CLASS2 | A | Q | SOT | A | TOP50 |
| CLASS2 | A | Q | SOT | Q | BOT50 |
| CLASS2 | A | Q | SOT | A | BOT50 |
| CLASS2 | A | A | SOT | A | TOP50 |
| CLASS1 | A | Q | SOT | Q | TOP50 |
| CLASS1 | Q | A | SOT | A | TOP50 |
| CLASS2 | A | Q | SOL | Q | BOT50 |
| CLASS2 | A | A | SOL | A | TOP50 |
| CLASS2 | A | Q | SOT | Q | BOT50 |
| CLASS1 | A | Q | SOL | A | BOT50 |
| CLASS1 | A | A | SOL | A | TOP50 |
| CLASS1 | A | A | SOL | A | TOP50 |
| CLASS1 | Q | A | SOL | A | TOP50 |
| CLASS1 | Q | Q | SOT | A | TOP50 |
| CLASS2 | Q | Q | SOL | Q | BOT50 |
| CLASS2 | Q | Q | SOL | Q | BOT50 |
| CLASS1 | Q | A | SOL | A | BOT50 |

*ENTEST:Q*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HIGHSCH** | **CLSRANK** | **ACT** | **MAJOR** | **ENTTEST** | **COLRANK** |
| CLASS1 | A | A | SOT | A | TOP50 |
| CLASS2 | Q | Q | SOL | Q | BOT50 |
| CLASS1 | A | A | SOT | A | TOP50 |
| CLASS2 | A | Q | SOT | A | TOP50 |
| CLASS2 | A | Q | SOT | Q | BOT50 |
| CLASS2 | A | Q | SOT | A | BOT50 |
| CLASS2 | A | A | SOT | A | TOP50 |
| CLASS1 | A | Q | SOT | Q | TOP50 |
| CLASS1 | Q | A | SOT | A | TOP50 |
| CLASS2 | A | Q | SOL | Q | BOT50 |
| CLASS2 | A | A | SOL | A | TOP50 |
| CLASS2 | A | Q | SOT | Q | BOT50 |
| CLASS1 | A | Q | SOL | A | BOT50 |
| CLASS1 | A | A | SOL | A | TOP50 |
| CLASS1 | A | A | SOL | A | TOP50 |
| CLASS1 | Q | A | SOL | A | TOP50 |
| CLASS1 | Q | Q | SOT | A | TOP50 |
| CLASS2 | Q | Q | SOL | Q | BOT50 |
| CLASS2 | Q | Q | SOL | Q | BOT50 |
| CLASS1 | Q | A | SOL | A | BOT50 |

*Predictability:*

|  |  |  |
| --- | --- | --- |
| ENTTEST | | |
| COLRANK | A | Q |
| TOP50 | 10 | 1 |
| BOT50 | 3 | 6 |

Predictability of using A implies TOP50 and Q implies BOT50

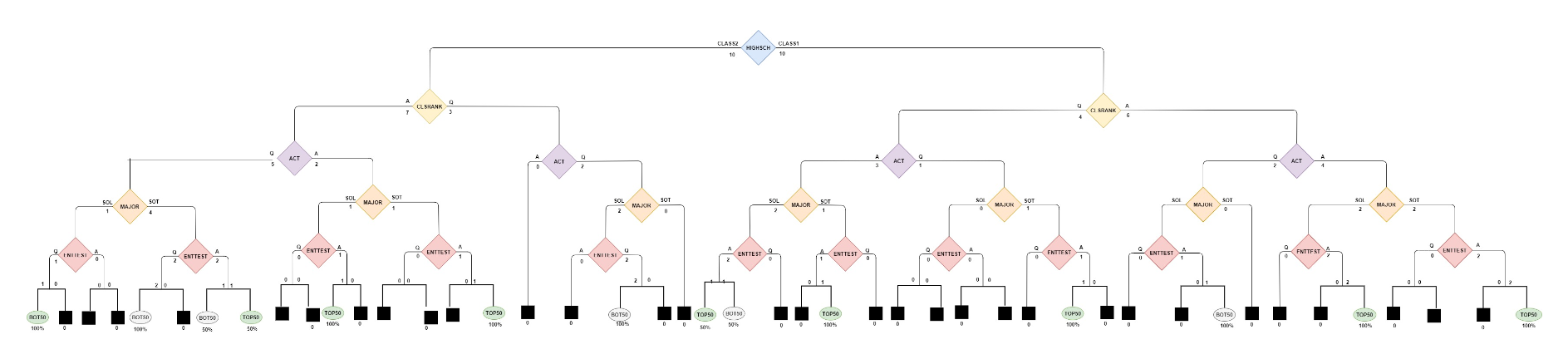
We would be

RIGHT: (10 + 6 ) / ( 10 + 6 + 3 + 1) : 16/20 times

WRONG: ( 3 + 1 ) / ( 10 + 6 + 3 + 1) : 4/20 times

*Decision Tree :*

Here, the order of attributes taken to build the tree is based on the availability of the attributes(WHEN AVAILABLE) – HIGHSCH was available first, so HIGHSCH is taken as root node and the rest of the attributes along the branches are chosen in the same order of availability too



Attached the screenshot here,to give a clear picture.



**3. In what ways did the second decision tree change the predictability of student outcomes over the first decision tree (and improve your decision making capability)?**

The first decision tree is built just based on two attributes whereas the second decision tree is built based on 5 attributes. The chances of classifying correctly and precisely the students to belong to TOP50 and BOT50 categories, is very high in the second model compared to that of the first model.

For example, as per decision tree1, students of class2, having ‘acceptable’ class rank , have 50/50 chances of belonging to TOP50 and BOT50 categories. Here, it just says 50% of students with acceptable class rank belong to TOP50 and the other 50% belong to BOT50. It does not say which 50% belong to each of these categories.

But the same case in decision tree2 explains which set of students belong to TOP50 and which set belong to BOT50 categories.

As per decision tree2,

7 students of class2 having acceptable class rank have acceptable ACT as well.

Out of these 7,

5 students have Major from SOT:

out of these 5, 3 have acceptable ENTTEST , out of these 3, 3 are classified to belong to TOP50 category and the other one is classified to belong to BOT50 category .

2 students have Major from SOL:

Out of these2, one student has acceptable ENTTEST and is classified to belong to TOP50 category whereas the other one has questionable ENTTEST and is classified to belong to BOT50 category.

This example clearly shows that the level of precision and quality of classification is very good with the decision tree2 when compared to the decision tree1.

**4. Using the Market Basket Analysis Technique develop a co-occurrence matrix of the courses that the students have registered for in a single semester. Based on this analysis and budget decision that you can only offer five of these courses each semester, which five courses would you offer together in the same semester and what other five courses would you offer together in the next semester.**

**ITM 531 540 561**

**ITM 528 531 561**

**ITM 531 561 570**

**ITM 540 561 578**

**ITM 521 540 555**

**ITM 561 548 578**

**ITM 531 561 578**

**ITM 531 561 570**

**ITM 528 548 578**

**ITM 540 561 570**

**ITM 528 531 561**

**ITM 528 548 578**

**ITM 540 521 555**

**ITM 540 570 555**

**ITM 570 528 561**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Courses | 578 | 570 | 561 | 555 | 548 | 540 | 531 | 528 | 521 |
| 578 | 5 | 0 | 3 | 0 | 3 | 1 | 1 | 2 | 0 |
| 570 | 0 | 5 | 4 | 1 | 0 | 2 | 2 | 1 | 0 |
| 561 | 3 | 3 | 10 | 0 | 1 | 3 | 6 | 3 | 0 |
| 555 | 0 | 1 | 0 | 3 | 0 | 3 | 0 | 0 | 1 |
| 548 | 3 | 0 | 1 | 0 | 3 | 0 | 0 | 2 | 0 |
| 540 | 1 | 2 | 3 | 3 | 0 | 6 | 1 | 0 | 2 |
| 531 | 1 | 2 | 5 | 0 | 0 | 1 | 6 | 2 | 0 |
| 528 | 2 | 1 | 3 | 0 | 2 | 0 | 2 | 5 | 0 |
| 521 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 2 |
| SUM | 10 | 9 | 19 | 5 | 6 | 12 | 12 | 10 | 3 |

Courses that can be offered together for the same semester:

From the matrix the top 5 courses which are enrolled by most students are :

ITM561,ITM540,ITM531,ITM528,ITM578

Hence, the list of 5 courses that can be suggested for the same semester are :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITM561 | ITM540 | ITM531 | ITM528 | ITM578 |

Courses that can be offered together for the next semester:

The remaining 4 courses -ITM570,ITM555,ITM548,ITM521can be suggested for the next semester.1 more course that can be suggested for the next semester will be the popular course-ITM561 since it has more enrollments.

Hence, the list of 5 courses that can be suggested for the next semester are :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITM570 | ITM555 | ITM548 | ITM521 | ITM561 |