Data Mining for Curriculum Redesign

Data: Courses.doc, Enrollment.csv
Tools Used: KNIME Analytics, Excel
Techniques Used: String distance, clustering, association rules for classification, frequent pattern matching

Kimberly Healy
healy.kim@gmx.us

Table of Contents

Section 1

Problem Definition

Over the past five years, the Department of Arts of the University of Diploma Printing in Romanigstan experienced stagnant enrollment. The Bachelor of Fine Arts (BFA) program has been selected to undergo review with the goal of attracting interest and boosting enrollment numbers in the program.

The current curriculum requires 39 credit hours for degree obtainment. Courses are broken down into three areas: foundation, core, and elective. The foundation and core areas have predetermined courses that are required for all students in the program. The foundation area requires two courses and the core area requires seven courses. The elective area does not have predetermined courses; the students are required to choose six out of the 33 elective course offerings. The curriculum courses are mainly art, history, language, politics, and literature related, but there are also a limited number of courses from other disciplines, including "Analytical Mechanics," "Computer Linear Algebra," "Behavioral Pharmacology" and "Cell Biology & Biochemistry".

Possible reasons for stagnant enrollment in the BFA program include shifting disciplinary approaches and changing student interests and demographics. The Committee Chair of the Department of Arts is interested in redesigning the BFA curriculum to better suit the evolving pedagogy and student environment.

Objectives of Data Mining Experiment

The objective of the data mining experiment is to provide actionable insights and advice for a curriculum redesign. Suggested redesign includes:

(1) reducing required credit hours from 39 to 33,
(2) reducing the number of offered electives, and
(3) defining concentration areas for elective courses.

This experiment will concentrate on solutions for (2) and (3) above. In order to accomplish this, the experiment must provide data-based evidence for which electives should be cut from the curriculum and what elective courses should be grouped together in concentration areas.

Section 2

Dataset Description

Two datasets are provided: the BFA Curriculum course offering and Student Enrollment by semester over the period 2000-2005.

BFA Curriculum dataset

The BFA Curriculum dataset is a document with a list of the offered courses for the BFA program. The dataset groups the course offerings into Foundation (2), Core (7), and Elective (33) areas. Every individual course offering is identified by a number (Course #) and name (Course Name).

Attribute	Variable Type	Unique instances	Missing instances
Course #	String, Integer	42	0
Course Name	String	42	0

Student Enrollment dataset

The Student Enrollment dataset is a CSV file that details enrollment in the program in the Fall, Spring, and Summer semesters from 2000-2005. The dataset includes student name, semester the course was taken (semester new), and the course name (coursename).

The following chart is a snapshot of the student enrollment dataset:

Attribute	Variable Type	Unique instances	Missing instances	Top occurrence	Top occurrence frequency	Bottom occurrence	Bottom occurrence frequency
Student name	String	448	0	Ed McMahon Harvey Glob	52	*[many]*	1
semester new	String, Integer	16	0	Spring 2002	486	Summer 2003	151
coursename	String	169	1	COMPUT LINEAR ALGEBRA	411	*[many]*	1

See **Appendix 1** for a summary of the Top 20 and Bottom 20 variables.

Problems with the raw datasets include the following:

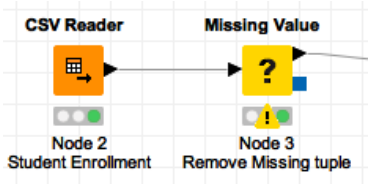
	Problem	Example	Preprocessing Solution
1	Missing data instances.	John Jakes, Summer 2003, *[Missing course name]*	Remove missing tuples.
2	Duplicate tuples.	Gray Davis, Spring 2003, COMPU LINEAR ALGEBRA appears in 5 different tuples	Remove duplicates.
3	Misspelling of course names from Curriculum dataset to Student Enrollment dataset.	"AMERICAN HEALTH POLICY" on the Curriculum dataset "AMERICAN HEALT POLICY" on the Student Enrollment dataset	Cluster similar course names then use the Curriculum dataset as a dictionary to replace

			misspelled course names in the Student Enrollment dataset.
4	Repetitive courses in the Curriculum dataset.	FRANCE AND THE EUROPEAN UNION is represented by two difference course numbers, ARTS 571 and ARTS 577	Remove ARTS 577 tuple.
5	Inconsistency of student name format in the Enrollment dataset.	Kofi A. Annan H. Normal Schwarzkopf Julie Christie 1995 Gov. Mario Cuomo Dr. Benjamin Spock Frank Sinatra Jr. Sark R. L. Stine QMarolyn Quayle Elvira (Cassandra Peterson) J.A.Jance state representative	Split student name cell and create more applicable attributes such as: First Name Middle Name Last Name Prefix/Suffix
6	Some course names in the student enrollment dataset are not courses offered in the BFA curriculum dataset.	"BRITAIN SINCE 1945" "1ST YR CLASSICAL CHIN II" "ECONOMETRIC FORECASTING"	Create two datasets: one dataset contains Enrollment data for courses within BFA Curriculum; one dataset contains the excess Enrollment data
7	No assignment of curriculum area (Core, Foundation, Elective) to Curriculum and Enrollment datasets	n/a	Assign area to tuples of the Curriculum dataset then join with Enrollment dataset.

Section 3

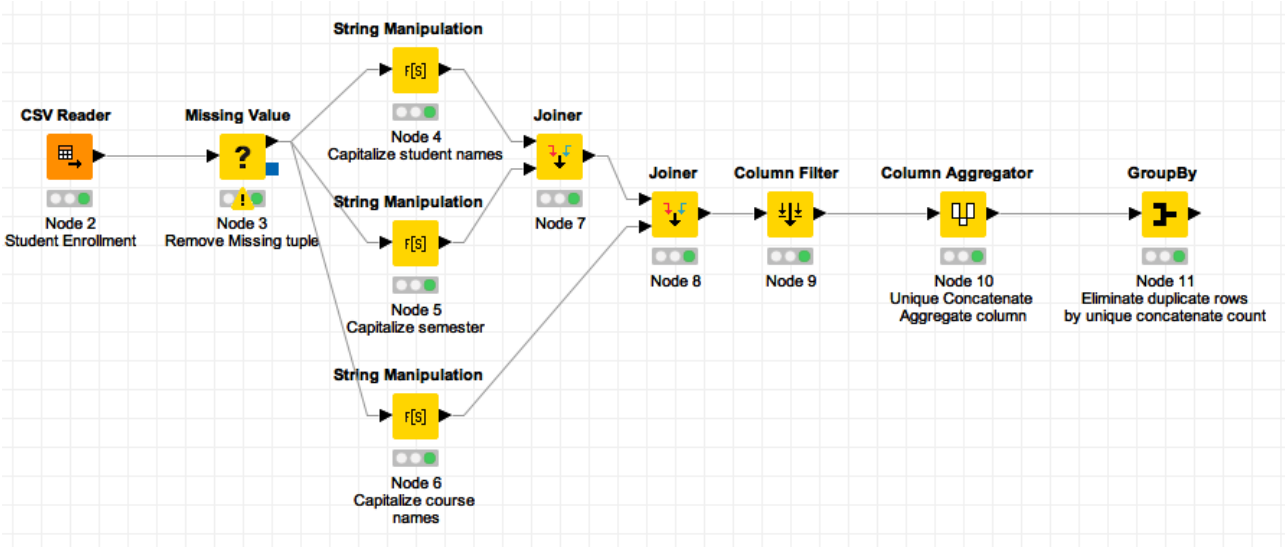
Data Preprocessing

Problem 1: Missing Values
Problem Analysis: There is only one coursename missing value in the Student Enrollment dataset. Since this instance is very small compared to the entire dataset, and no external sources can verify the missing data, the entire tuple is removed.
Solution 1: KNIME Workflow



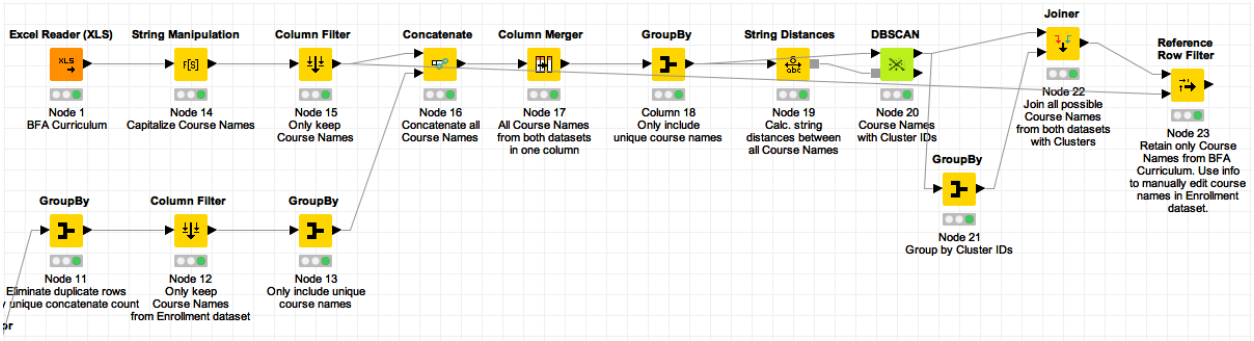
Total Rows Node 2: 4900
Total Rows Node 3: 4899

Problem 2: Duplicate tuples
Problem Analysis: There are a reported 1249 tuples containing the same information. To fix this problem, the attributes Student name, semester new, and coursename are first normalized (capitalized) then aggregated (concatenated) into one column, Unique Concatenate. The tuples are then grouped by the Unique Concatenate count. Any count greater than 1 means that the tuple has an exact duplicate. The remaining table only contains unique tuples.
Solution 2: KNIME Workflow

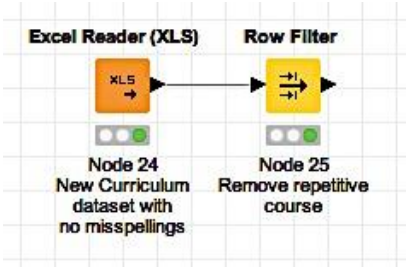


Total Rows Node 10: 4899
Total Rows Node 11: 3650

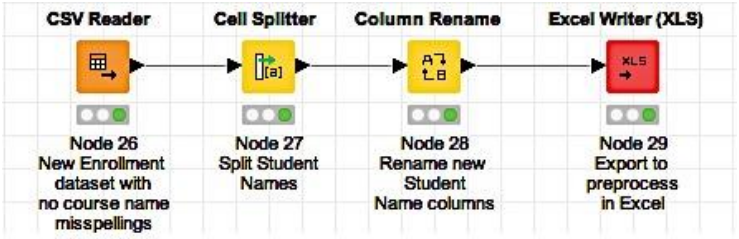
Problem 3: Misspelling of course names
Problem Analysis: The Student Enrollment dataset contains misspellings of the course names listed in the Curriculum dataset. The courses represent the same real-world entity but are misspelled or have mismatching punctuation. To fix this problem, similar course names will be clustered together using string distances. The clusters will reduce the time spent on manually evaluating all course names in the Student Enrollment dataset to the correct spelling in the Curriculum dataset. With the knowledge gained from clustering, the misspellings will be corrected in Excel using Find>Replace tool. See **Appendix 2** for the corrected names in the BFA Curriculum dataset and **Appendix 3** for the corrected names in the Enrollment dataset.
Solution 3: KNIME Workflow



Problem 4: Repetitive courses in the Curriculum dataset.
Problem Analysis: The Curriculum dataset contains two tuples that represent the same real-world entity: the FRANCE AND THE EUROPEAN UNION course. This course is assigned two different course numbers, ARTS 571 and ARTS 577. To fix this problem, remove ARTS 577 repetitive tuple. The resulting dataset is the finalized Curriculum dataset.
Solution 4: KNIME Workflow



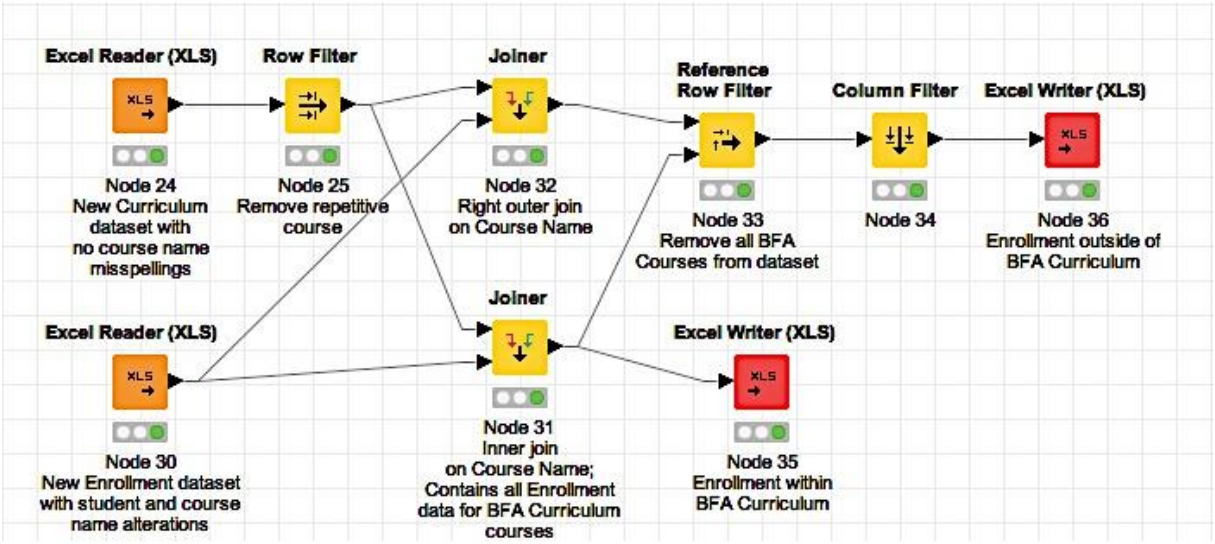
Problem 5: Inconsistency of student names in the Enrollment dataset
Problem Analysis: Student names are not entered in a consistent manner into the Enrollment data. To reduce the manual work to solve this problem, split the names in the Student Name column by a " " space delimiter then process further in Excel.
Solution 5: KNIME Workflow and Excel Results



FIRST NAM	LAST NAME	MIDDLE NA	PREFIX/SUF	SEMESTER	COURSE NAME
ABELLA	ABZUG			FALL 2003	A WORLD AT WAR
ABELLA	ABZUG			FALL 2003	COMPUTER LINEAR ALGEBRA
ABELLA	ABZUG			FALL 2003	ENVIRONMENTAL CASE STUDIES
ABELLA	ABZUG			FALL 2004	ANALYTICAL MECHANICS
ABELLA	ABZUG			FALL 2004	BEHAVIORAL PHARMACOLOGY
ABELLA	ABZUG			FALL 2004	EXPERIMENTAL WRITING SEMINAR
ABELLA	ABZUG			FALL 2005	COMPARATIVE POLITICS
ABELLA	ABZUG			FALL 2005	FOOD/FEAST ARCH OF TABLE

Total row count remains at 3650.

Problem 6: Course names not offered in BFA curriculum appear in Enrollment dataset
Problem Analysis: The Enrollment dataset contains students taking courses both within the BFA curriculum and outside of the BFA curriculum. For future reference, create two datasets: one dataset contains Enrollment data for courses within the BFA Curriculum and one dataset contains Enrollment data for courses outside of the BFA Curriculum. The resulting 3 datasets are as follows: Total Enrollment, Enrollment Outside BFA Curriculum, and Enrollment Within BFA Curriculum.
Solution 6: KNIME Workflow



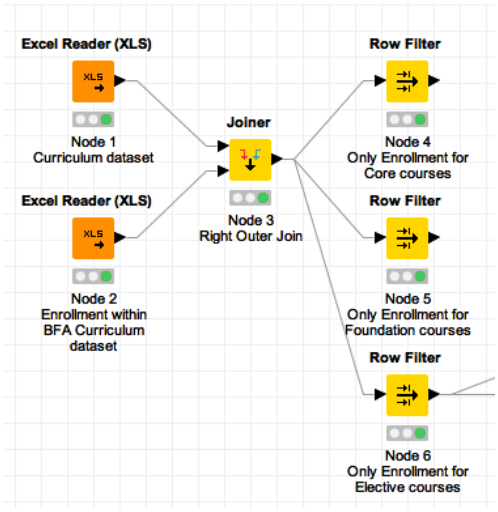
Problem 7: No assignment of curriculum area (Core, Foundation, Elective) to Enrollment Within BFA Curriculum dataset
Problem Analysis: The problems this project attempts to solve focuses on the elective course offerings. It will be helpful in the future to isolate enrollment in each of the three areas (Core, Foundation, Elective). To do so, create a new column in the Curriculum dataset called "AREA" and assign each course the appropriate area. Then, join the Curriculum dataset with the Enrollment Within BFA Curriculum dataset

on Course Name so each Enrollment tuple has the correct corresponding Area assignment. Filter the Enrollment Within BFA Curriculum dataset by Area. The resulting 3 datasets are as follows: Enrollment of Core courses, Enrollment of Foundation courses, and Enrollment of Elective courses.
Solution 7: Excel Area assignment and KNIME Workflow

Excel Area assignment to the Curriculum dataset;

	A	B	C
1	COURSE #	COURSE NAME	AREA
2	ARTS 400	EXPERIMENTAL W	FOUNDATION
3	ARTS 401	ART: ANCIENT TO	FOUNDATION
4	ARTS 465	ENVIRONMENTAL	CORE
5	ARTS 486	COMPUTER LINEA	CORE
6	ARTS 512	ANALYTICAL MECI	CORE
7	ARTS 514	A WORLD AT WAF	CORE
8	ARTS 516	BEHAVIORAL PHA	CORE
9	ARTS 518	CONTEMPORARY	CORE
10	ARTS 520	FOOD/FEAST ARCI	CORE
11	ARTS 488	DEVIL'S PACT LITE	ELECTIVE
12	ARTS 541	AMERICAN SOCIA	ELECTIVE
13	ARTS 543	ART AND RELIGIOI	ELECTIVE
14	ARTS 491	CONTEMPORARY	ELECTIVE
15	ARTS 492	AFRICAN AMERIC/	ELECTIVE
16	ARTS 493	AMERICAN HEALT	ELECTIVE

KNIME Area assignment to Enrollment Within BFA Curriculum dataset and filtered datasets by Area



Dataset Summary After Preprocessing

Dataset	Total Rows
Curriculum	41
Enrollment Outside BFA Curriculum	1267
Enrollment Within BFA Curriculum	2383
Enrollment of Core courses	1461
Enrollment of Foundation courses	205
Enrollment of Elective courses	717
Total Enrollment	3650

Section 4

Mining Goals

The goals of this mining project include the following:
- Find data-based evidence to reduce the number of offered electives.
- Find data-based evidence to define concentration areas for elective courses.

Mining

Task 1: Find the most and least popular Elective courses.

The purpose of this task is to find which elective courses within the BFA Curriculum were the most and least popular over the semesters included in the Enrollment dataset. By gaining this knowledge, we can have an understanding of the unpopular elective courses that can be removed from the course offering. The Enrollment of Elective courses dataset is analyzed. It has a total of 717 tuples.

The following are the top 15 Elective courses:

Course #	Course Name	Frequency Count	Approx. Relative Frequency	Popularity Rank
ARTS 493	AMERICAN HEALTH POLICY	134	18.69%	1
ARTS 569	CELLULAR BIOLOGY AND BIOCHEMISTRY	81	11.3%	2
ARTS 587	ELEMENTARY ARABIC II	64	8.93%	3
ARTS 547	COMMUNICATIONS INTERNSHIP	45	6.28%	4
ARTS 491	CONTEMPORARY POLITICAL THOUGHT	44	6.14%	5
ARTS 549	FRESHWATER ECOLOGY	41	5.72%	6
ARTS 557	19TH CENTURY BRITISH LITERATURE	39	5.44%	7
ARTS 555	BECOMING HUMAN	32	4.46%	8
ARTS 545	20TH CENTURY RUSSIAN LITERATURE: FICTION AND REALITY	31	4.32%	9
ARTS 494	BUSINESS GERMAN: A MICRO PERSPECTIVE	25	3.49%	10
ARTS 581	COMPARATIVE POLITICS	23	3.21%	11
ARTS 492	AFRICAN AMERICAN LITERATURE	20	2.80%	12
ARTS 571	FRANCE AND THE EUROPEAN UNION	16	2.23%	13
ARTS 565	ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL	14	1.95%	14
ARTS 488	DEVIL'S PACT LITERATURE/FILM	12	1.67%	15

The following are the bottom 12 Elective courses:

Course #	Course Name	Frequency Count	Approx. Relative Frequency	Popularity Rank
ARTS 551	AESTHETICS	12	1.67%	16
ARTS 583	BRITISH POETRY 1660-1914	12	1.67%	17
ARTS 484	EUROPE IN A WIDER WORLD	11	1.53%	18
ARTS 543	ART AND RELIGION	10	1.39%	19
ARTS 575	EARLY MESOPOTAMIAN HISTORY/SOCIETY	10	1.39%	20
ARTS 561	AUGUSTAN CULTURAL REVOLUTION	8	1.12%	21
ARTS 495	COMMUNICATIONS AND THE PRESIDENCY	7	0.98%	22
ARTS 553	FRENCH THOUGHT SINCE 1945	7	0.98%	23
ARTS 497	CONTEMPORARY ART: 1945 TO PRESENT	6	0.84%	24
ARTS 567	ELEMENTARY GERMAN 1	6	0.84%	25
ARTS 573	ANALYZING THE POLITICAL WORLD	6	0.84%	26
ARTS 585	CONTEMPORARY SOCIO THEORY	1	0.14%	27

Task 2: Find elective courses that are associated with the top 3 most frequently taken elective courses.

The purpose of this mining task is to get an idea of which elective courses are taken with the most popular elective courses. By gaining this knowledge, we can strengthen the divide between the most and least popular elective courses.

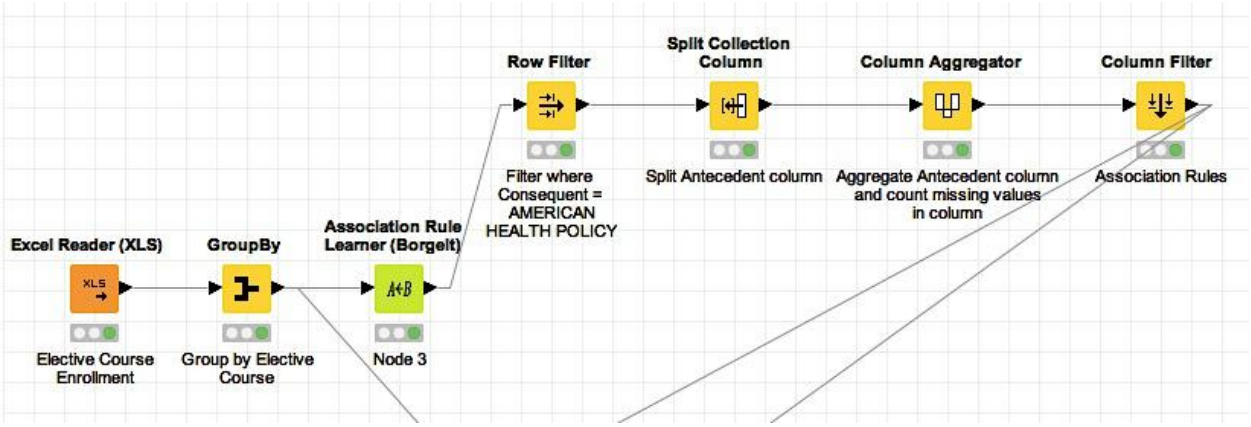
Association rules for classification is a technique where the miner defines the Consequent to be mined. In this case, the Consequent elective courses are the top 3 most frequently taken elective courses. Those courses are found in Task 1 above:

- AMERICAN HEALTH POLICY
- CELLULAR BIOLOGY AND BIOCHEMISTRY
- ELEMENTARY ARABIC II

The courses in the Antecedent will indicate which courses are most commonly associated with the 3 most popular courses listed above. The rules are ranked by the calculation of their lift, a measurement that indicates the ability of the rule to tell the miner an important association as opposed to a random association. Lifts greater than 1 are considered actionable association rules.

Prepare for association rule for classification mining:

Association Rule for Classification KNIME Workflow



Mining parameters

Minimum set size: 2
Minimum support: 10%
Minimum rule confidence: 10

Top 10 Association Rules of Consequent AMERICAN HEALTH POLICY by most actionable (highest Lift). Popularity rank of each course in the Antecedent is given in parenthesis:

Antecedent	Consequent	Support Count	Confidence %	Lift
FRESHWATER ECOLOGY (6), CONTEMPORARY POLITICAL THOUGHT (5)	AMERICAN HEALTH POLICY	8	72.7	1.966
FRESHWATER ECOLOGY (6)	AMERICAN HEALTH POLICY	28	68.3	1.846
DEVIL'S PACT LITERATURE/FILM (15)	AMERICAN HEALTH POLICY	7	63.6	1.72
FRESHWATER ECOLOGY (6), ELEMENTARY ARABIC II (3)	AMERICAN HEALTH POLICY	6	54.5	1.475

BUSINESS GERMAN: A MICRO PERSPECTIVE (10)	AMERICAN HEALTH POLICY	13	52	1.406
COMMUNICATIONS INTERNSHIP (4)	AMERICAN HEALTH POLICY	23	51.1	1.382
COMMUNICATIONS INTERNSHIP (4), CELLULAR BIOLOGY AND BIOCHEMISTRY (2)	AMERICAN HEALTH POLICY	5	50	1.352
COMMUNICATIONS INTERNSHIP (4), ELEMENTARY ARABIC II (3)	AMERICAN HEALTH POLICY	5	50	1.352
CONTEMPORARY POLITICAL THOUGHT (5)	AMERICAN HEALTH POLICY	19	43.2	1.167
CONTEMPORARY POLITICAL THOUGHT (5), CELLULAR BIOLOGY AND BIOCHEMISTRY (2)	AMERICAN HEALTH POLICY	4	40	1.081

Top 5 Association Rules of Consequent CELLULAR BIOLOGY AND BIOCHEMISTRY by most actionable (highest Lift). Popularity rank of each course in the Antecedent is given in parenthesis:

Antecedent	Consequent	Support Count	Confidence %	Lift
DEVIL'S PACT LITERATURE/FILM (15)	CELLULAR BIOLOGY AND BIOCHEMISTRY	4	36.4	1.397
AESTHETICS (16)	CELLULAR BIOLOGY AND BIOCHEMISTRY	4	33.3	1.281
ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL (14)	CELLULAR BIOLOGY AND BIOCHEMISTRY	4	28.6	1.098
EUROPE IN A WIDER WORLD (18)	CELLULAR BIOLOGY AND BIOCHEMISTRY	3	27.3	1.048
COMPARATIVE POLITICS (11)	CELLULAR BIOLOGY AND BIOCHEMISTRY	6	26.1	1.002

Top 10 Association Rules of Consequent ELEMENTARY ARABIC II by most actionable (highest Lift). Popularity rank of each course in the Antecedent is given in parenthesis:

Antecedent	Consequent	Support Count	Confidence %	Lift
ART AND RELIGION (19)	ELEMENTARY ARABIC II	5	50	2.281
FRANCE AND THE EUROPEAN UNION (13)	ELEMENTARY ARABIC II	6	37.5	1.711
BECOMING HUMAN (8)	ELEMENTARY ARABIC II	12	37.5	1.711
EUROPE IN A WIDER WORLD (18)	ELEMENTARY ARABIC II	4	36.4	1.659
BRITISH POETRY 1660-1914 (17)	ELEMENTARY ARABIC II	4	33.3	1.521
BECOMING HUMAN (8), AMERICAN HEALTH POLICY (1)	ELEMENTARY ARABIC II	4	33.3	1.521
AFRICAN AMERICAN LITERATURE (12)	ELEMENTARY ARABIC II	6	30	1.369
FRESHWATER ECOLOGY (6)	ELEMENTARY ARABIC II	11	26.8	1.224
COMPARATIVE POLITICS (11)	ELEMENTARY ARABIC II	6	26.1	1.19
AESTHETICS (16)	ELEMENTARY ARABIC II	3	25	1.141

Task 3: Group together elective courses frequently taken together using frequent pattern matching.

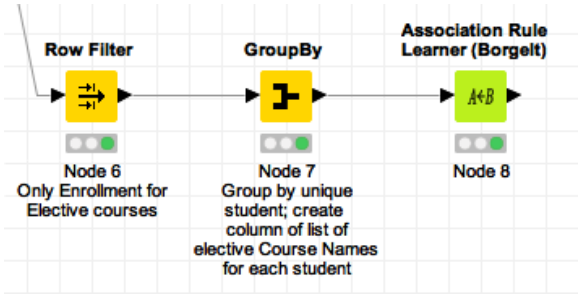
The purpose of this mining task is to get an idea of what elective courses are frequently taken together. By gaining this knowledge, we can provide evidence for groups of courses that can lead to a concentration area.

Association rules derived from frequent patterns in the data are found. Multiple combinations of minimum support and minimum confidence (the mining parameters) are used. The support indicates the statistical significance of the association rule in the database. The confidence indicates the degree of correlation in a database between the anecdote and consequent. Sets of two and three course offerings are found. The rules are ranked by the calculation of their lift, a measurement that indicates the ability of the rule to tell the miner an important association as opposed to a random association. Lifts greater than 1 are considered actionable association rules.

Prepare for association rule mining:

Use the Enrollment of Elective courses dataset. Group by unique student name and create a column with an aggregated list of each student's elective choices. The following is the KNIME Workflow and result of the GroupBy node. The List(COURSE NAME) column will be used for association rule mining.

Association Rule KNIME Workflow



GroupBy result:

Table "default" - Rows: 292 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	FIRST ...	LAST ...	MIDDL...	PREFI...	List(COURSE NAME)	List(COURSE #)
Row0	ABELLA	ABZUG	?	?	[20TH CENTURY RUSSIAN LITER...	[ARTS 545 ,ARTS 581]
Row1	AL	GORE	?	?	[AMERICAN HEALTH POLICY,20...	[ARTS 493,ARTS 545 ,ARTS 549,...]
Row2	AL	HIRT	?	?	[CELLULAR BIOLOGY AND BIOCH...	[ARTS 569]
Row3	AL	ROKER	?	?	[AMERICAN HEALTH POLICY]	[ARTS 493]
Row4	ALAN	BATES	?	?	[AMERICAN HEALTH POLICY,CO...	[ARTS 493,ARTS 547,ARTS 555,...]
Row5	ALEC	BALDWIN	?	?	[CONTEMPORARY POLITICAL TH...	[ARTS 491,ARTS 493,ARTS 493]
Row6	ANDY	GARCIA	?	?	[AFRICAN AMERICAN LITERATU...	[ARTS 492]
Row7	ANGELA	LANSBURY	?	?	[AMERICAN HEALTH POLICY,CO...	[ARTS 493,ARTS 497,ARTS 545]
Row8	ANGELICA	HUSTON	?	?	[19TH CENTURY BRITISH LITER...	[ARTS 557]
Row9	ANN	LANDERS	?	?	[AMERICAN HEALTH POLICY]	[ARTS 493]
Row10	ANN	RICHARDS	W.	GOV.	[AMERICAN HEALTH POLICY]	[ARTS 493]
Row11	ANNE	PERRY	?	?	[EARLY MESOPOTAMIAN HISTOR...	[ARTS 575]
Row12	ANNETTE	FUNICELLO	?	?	[CONTEMPORARY POLITICAL TH...	[ARTS 491,ARTS 569 ,ARTS 581 ,...
Row13	ANNIE	LENNOX	?	?	[COMMUNICATIONS INTERNSHIP]	[ARTS 547]
Row14	ANTHONY	HOPKINS	?	?	[DEVIL'S PACT LITERATURE/FIL...	[ARTS 488 ,ARTS 493]
Row15	ARCHER	MAYOR	?	?	[DEVIL'S PACT LITERATURE/FIL...	[ARTS 488 ,ARTS 494,ARTS 547,...]
Row16	ARMAND	ASSANTE	?	?	[CONTEMPORARY POLITICAL TH...	[ARTS 491,ARTS 492 ,ARTS 545 ,...
Row17	ARMORY	HOUGHTON	?	JR.	[AFRICAN AMERICAN LITERATU...	[ARTS 492 ,ARTS 565,ARTS 569 ,...
Row18	ARSENIO	HALL	?	?	[ANALYZING THE POLITICAL WO...	[ARTS 573,ARTS 587]
Row19	ARTE	JOHNSON	?	?	[AMERICAN HEALTH POLICY,CO...	[ARTS 493,ARTS 547]
Row20	BARBARA	BRADFORD	TAYLOR	?	[AMERICAN HEALTH POLICY,BE...	[ARTS 493,ARTS 555,ARTS 565]

Find association rules amongst Elective courses:

Mining parameters – 1

Minimum set size: 2
Minimum support: 10%
Minimum rule confidence: 10

Top 10 Association Rules by most actionable (highest Lift):

Antecedent	Consequent	Support Count	Confidence %	Lift
20TH CENTURY RUSSIAN LITERATURE: FICTION AND RELAITY	19TH CENTURY BRITISH LITERATURE	13	41.9	3.14
19TH CENTURY BRITISH LITERATURE	20TH CENTURY RUSSIAN LITERATURE: FICTION AND RELAITY	13	33.3	3.14
BECOMING HUMAN	FRANCE AND THE EUROPEAN UNION	5	15.6	2.852
19TH CENTURY BRITISH LITERATURE	BRITISH POETRY 1660-1914	4	10.3	2.496
CONTEMPORARY POLITICAL THOUGHT	ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL	5	11.4	2.37
20TH CENTURY RUSSIAN LITERATURE: FICTION AND RELAITY	AFRICAN AMERICAN LITERATURE	5	16.1	2.355
CONTEMPORARY POLITICAL THOUGHT	AFRICAN AMERICAN LITERATURE	6	13.6	1.991
FRESHWATER ECOLOGY	AMERICAN HEALTH POLICY	28	68.3	1.846
AMERICAN HEALTH POLICY	FRESHWATER ECOLOGY	28	25.9	1.846
FRESHWATER ECOLOGY	CONTEMPORARY POLITICAL THOUGHT	11	26.8	1.78

Mining parameters - 2

Minimum set size: 3
Minimum support: 10%
Minimum rule confidence: 10

Top 10 Association Rules by most actionable (highest Lift):

Antecedent	Consequent	Support Count	Confidence %	Lift
CONTEMPORARY POLITICAL THOUGHT CELLULAR BIOLOGY AND BIOCHEMISTRY	ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL	3	30	6.257
COMMUNICATIONS INTERNSHIP ELEMENTARY ARABIC II	ART AND RELIGION	2	20	5.84
COMMUNICATIONS INTERNSHIP ELEMENTARY ARABIC II	BRITISH POETRY 1660-1914	2	20	4.867
BECOMING HUMAN ELEMENTARY ARABIC II	FRANCE AND THE EUROPEAN UNION	3	25	4.562
BECOMING HUMAN ELEMENTARY ARABIC II	EUROPE IN A WIDER WORLD	2	16.7	4.424

CONTEMPORARY POLITICAL THOUGHT CELLULAR BIOLOGY AND BIOCHEMISTRY	COMMUNICATIONS AND THE PRESIDENCY	1	10	4.171
19TH CENTURY BRITISH LITERATURE AMERICAN HEALTH POLICY	FRENCH THOUGHT SINCE 1945	1	10	4.171
CELLULAR BIOLOGY AND BIOCHEMISTRY AMERICAN HEALTH POLICY	AUGUSTAN CULTURAL REVOLUTION	2	10.5	3.842
20TH CENTURY RUSSIAN LITERATURE: FICTION AND REALITY 19TH CENTURY BRITISH LITERATURE	BRITISH POETRY 1660-1914	2	15.4	3.744
COMMUNICATIONS INTERNSHIP AMERICAN HEALTH POLICY	DEVIL'S PACT LITERATURE/FILM	3	13	3.462

Mining parameters - 3

Minimum set size: 2
Minimum support: 15%
Minimum rule confidence: 15

Top 10 Association Rules by most actionable (highest Lift):

Antecedent	Consequent	Support Count	Confidence %	Lift
AMERICAN HEALTH POLICY	FRESHWATER ECOLOGY	28	25.9	1.846
CONTEMPORARY POLITICAL THOUGHT	FRESHWATER ECOLOGY	11	25	1.78
ELEMENTARY ARABIC II	BECOMING HUMAN	12	18.8	1.711
COMMUNICATIONS INTERNSHIP	BECOMING HUMAN	7	15.6	1.419
COMMUNICATIONS INTERNSHIP	AMERICAN HEALTH POLICY	23	51.1	1.382
AMERICAN HEALTH POLICY	COMMUNICATIONS INTERNSHIP	23	21.3	1.382
COMMUNICATIONS INTERNSHIP	FRESHWATER ECOLOGY	8	17.8	1.266
ELEMENTARY ARABIC II	FRESHWATER ECOLOGY	11	17.2	1.224
CONTEMPORARY POLITICAL THOUGHT	COMMUNICATIONS INTERNSHIP	8	18.2	1.18
COMMUNICATIONS INTERNSHIP	CONTEMPORARY POLITICAL THOUGHT	8	17.8	1.18

Mining parameters - 4

Minimum set size: 3
Minimum support: 15%
Minimum rule confidence: 15

Top 10 Association Rules by most actionable (highest Lift):

Antecedent	Consequent	Support Count	Confidence %	Lift
CONTEMPORARY POLITICAL THOUGHT, AMERICAN HEALTH POLICY	FRESHWATER ECOLOGY	8	42.1	2.999

FRESHWATER ECOLOGY, AMERICAN HEALTH POLICY	BUSINESS GERMAN: A MICRO PERSPECTIVE	6	21.4	2.503
FRESHWATER ECOLOGY, AMERICAN HEALTH POLICY	CONTEMPORARY POLITICAL THOUGHT	8	28.6	1.896
ELEMENTARY ARABIC II, AMERICAN HEALTH POLICY	FRESHWATER ECOLOGY	6	26.1	1.858
CONTEMPORARY POLITICAL THOUGHT, AMERICAN HEALTH POLICY	COMMUNICATIONS INTERNSHIP	5	26.3	1.708
CELLULAR BIOLOGY AND BIOCHEMISTRY, AMERICAN HEALTH POLICY	COMMUNICATIONS INTERNSHIP	5	26.3	1.708
COMMUNICATIONS INTERNSHIP, AMERICAN HEALTH POLICY	BECOMING HUMAN	4	17.4	1.587
ELEMENTARY ARABIC II, AMERICAN HEALTH POLICY	BECOMING HUMAN	4	17.4	1.587
COMMUNICATIONS INTERNSHIP, AMERICAN HEALTH POLICY	FRESHWATER ECOLOGY	5	21.7	1.548
COMMUNICATIONS INTERNSHIP, AMERICAN HEALTH POLICY	CONTEMPORARY POLITICAL THOUGHT	5	21.7	1.443

Section 5

Evaluation of Experiments

Goal: Find data-based evidence to reduce the number of offered electives.

By combining the results from Task 1 and Task 2 in Section 4, we can strengthen an argument for which elective courses should be retained and which should be removed from offering. The courses at risk of removal are ranked in the bottom 12 least popular elective courses. The following is a summary of the results:

- All courses associated with AMERICAN HEALTH POLICY are ranked in the top 15 most popular elective courses and are retained.
- The following courses are ranked in the 12 least popular elective courses but are associated with both CELLULAR BIOLOGY AND BIOCHEMISTRY and ELEMENTARY ARABIC II: AESTHETICS and EUROPE IN A WIDER WORLD. This is strong evidence to retain both courses.
- The following courses are ranked in the 12 least popular elective courses but are associated with ELEMENTARY ARABIC II: BRITISH POETRY 1660-1914 and ART AND RELIGION. This is evidence to retain both courses.

The following tables contain a list of the 19 retained and 8 removed offered elective courses.

COURSES RETAINED IN OFFERING	
Course #	**Course Name**
ARTS 493	AMERICAN HEALTH POLICY
ARTS 569	CELLULAR BIOLOGY AND BIOCHEMISTRY
ARTS 587	ELEMENTARY ARABIC II
ARTS 547	COMMUNICATIONS INTERNSHIP
ARTS 491	CONTEMPORARY POLITICAL THOUGHT
ARTS 549	FRESHWATER ECOLOGY
ARTS 557	19TH CENTURY BRITISH LITERATURE
ARTS 555	BECOMING HUMAN
ARTS 545	20TH CENTURY RUSSIAN LITERATURE: FICTION AND REALITY
ARTS 494	BUSINESS GERMAN: A MICRO PERSPECTIVE
ARTS 581	COMPARATIVE POLITICS
ARTS 492	AFRICAN AMERICAN LITERATURE
ARTS 571	FRANCE AND THE EUROPEAN UNION
ARTS 565	ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL
ARTS 488	DEVIL'S PACT LITERATURE/FILM
ARTS 551	AESTHETICS
ARTS 583	BRITISH POETRY 1660-1914
ARTS 484	EUROPE IN A WIDER WORLD
ARTS 543	ART AND RELIGION

COURSES REMOVED FROM OFFERING	
Course #	**Course Name**
ARTS 575	EARLY MESOPOTAMIAN HISTORY/SOCIETY
ARTS 561	AUGUSTAN CULTURAL REVOLUTION
ARTS 495	COMMUNICATIONS AND THE PRESIDENCY
ARTS 553	FRENCH THOUGHT SINCE 1945
ARTS 497	CONTEMPORARY ART: 1945 TO PRESENT
ARTS 567	ELEMENTARY GERMAN 1
ARTS 573	ANALYZING THE POLITICAL WORLD
ARTS 585	CONTEMPORARY SOCIO THEORY

Goal: Find data-based evidence to define concentration areas for elective courses.

By evaluating the results from the various minings performed in Task 3 in Section 4, we can strengthen an argument for concentration areas for the elective courses. The following tables contain noticeable groupings from each mining parameter.

Groupings from Mining Parameters - 1		
20TH CENTURY RUSSIAN LITERATURE: FICTION AND RELAITY	19TH CENTURY BRITISH LITERATURE	AFRICAN AMERICAN LITERATURE
BECOMING HUMAN	FRANCE THE THE EUROPEAN UNION	-
19TH CENTURY BRITISH LITERATURE	BRITISH POETRY 1660-1914	-
CONTEMPORARY POLITICAL THOUGHT	ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL	AFRICAN AMERICAN LITERATURE
FRESHWATER ECOLOGY	AMERICAN HEALTH POLICY	CONTEMPORARY POLITICAL THOUGHT

Groupings from Mining Parameters - 2			
20TH CENTURY RUSSIAN LITERATURE: FICTION AND RELAITY	19TH CENTURY BRITISH LITERATURE	BRITISH POETRY 1660-1914	FRENCH THOUGHT SINCE 1945
BECOMING HUMAN	ELEMENTARY ARABIC II	FRANCE THE THE EUROPEAN UNION	EUROPE IN A WIDER WORLD
COMMUNICATIONS INTERNSHIP	ELEMENTARY ARABIC II	ART AND RELIGION	BRITISH POETRY 1660-1914
CELLULAR BIOLOGY AND BIOCHEMISTRY	ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL	CONTEMPORARY POLITICAL THOUGHT	COMMUNICATIONS AND THE PRESIDENCY
AMERICAN HEALTH POLICY	CELLULAR BIOLOGY AND BIOCHEMISTRY	19TH CENTURY BRITISH LITERATURE	FRENCH THOUGHT SINCE 1945

Groupings from Mining Parameters - 3			
AMERICAN HEALTH POLICY	FRESHWATER ECOLOGY	-	-
CONTEMPORARY POLITICAL THOUGHT	FRESHWATER ECOLOGY	COMMUNICATIONS INTERNSHIP	-
COMMUNICATIONS INTERNSHIP	FRESHWATER ECOLOGY	ELEMENTARY ARABIC II	-
COMMUNICATIONS INTERNSHIP	AMERICAN HEALTH POLICY	BECOMING HUMAN	ELEMENTARY ARABIC II

Groupings from Mining Parameters - 4				
BUSINESS GERMAN: A MICRO PERSPECTIVE	AMERICAN HEALTH POLICY	FRESHWATER ECOLOGY	CONTEMPORARY POLITICAL THOUGHT	-
COMMUNICATIONS INTERNSHIP	AMERICAN HEALTH POLICY	CELLULAR BIOLOGY AND BIOCHEMISTRY	CONTEMPORARY POLITICAL THOUGHT	FRESHWATER ECOLOGY
BECOMING HUMAN	AMERICAN HEALTH POLICY	COMMUNICATIONS INTERNSHIP	ELEMENTARY ARABIC II	-

There is evidence from groupings such as [CELLULAR BIOLOGY AND BIOCHEMISTRY, ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL, CONTEMPORARY POLITICAL THOUGHT, and COMMUNICATIONS AND THE PRESIDENCY], [AMERICAN HEALTH POLICY, CELLULAR BIOLOGY AND BIOCHEMISTRY, 19TH CENTURY BRITISH LITERATURE, and FRENCH THOUGHT SINCE 1945], and [CONTEMPORARY POLITICAL THOUGHT and FRESHWATER ECOLOGY] that those students interested in health studies are also interested in political studies. Given that the students must select a minimum amount of courses from each concentration areas, the health studies and political studies areas should be separated so that the students can choose between them at will. The following tables contain courses that define the health studies and political studies concentration areas.

HEALTH STUDIES	
Course #	**Course Name**
ARTS 493	AMERICAN HEALTH POLICY
ARTS 569	CELLULAR BIOLOGY AND BIOCHEMISTRY
ARTS 549	FRESHWATER ECOLOGY
ARTS 565	ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL

POLITICAL STUDIES	
Course #	**Course Name**
ARTS 491	CONTEMPORARY POLITICAL THOUGHT
ARTS 494	BUSINESS GERMAN: A MICRO PERSPECTIVE
ARTS 581	COMPARATIVE POLITICS
ARTS 571	FRANCE AND THE EUROPEAN UNION
ARTS 484	EUROPE IN A WIDER WORLD

There is strong evidence from groupings such as [20TH CENTURY RUSSIAN LITERATURE: FICTION AND RELAITY, 19TH CENTURY BRITISH LITERATURE, BRITISH POETRY 1660-1914, and FRENCH THOUGHT SINCE 1945] and [20TH CENTURY RUSSIAN LITERATURE: FICTION AND RELAITY, 19TH CENTURY BRITISH LITERATURE, and AFRICAN AMERICAN LITERATURE] that the students who are interested in literature take many of the literature classes together. Thus, a literature studies concentration area should be formed.

LITERATURE STUDIES	
ARTS 557	19TH CENTURY BRITISH LITERATURE
ARTS 545	20TH CENTURY RUSSIAN LITERATURE: FICTION AND REALITY
ARTS 492	AFRICAN AMERICAN LITERATURE
ARTS 488	DEVIL'S PACT LITERATURE/FILM
ARTS 583	BRITISH POETRY 1660-1914

Lastly, groupings such as [COMMUNICATIONS INTERNSHIP, ELEMENTARY ARABIC II, and FRESHWATER ECOLOGY], [COMMUNICATIONS INTERNSHIP, ELEMENTARY ARABIC II, AMERICAN HEALTH POLICY, BECOMING HUMAN] and [COMMUNICATIONS INTERNSHIP, ELEMENTARY ARABIC II, ART AND RELIGION, and BRITISH POETRY 1660-1914] provide evidence that those students interested in communication and language studies are also interested in health studies and literature studies. Given that the students must select a minimum amount of courses from each concentration areas, a communications studies concentration area should be separated from the health studies and literature studies so that the students can choose between them at will.

COMMUNICATION STUDIES	
Course #	**Course Name**
ARTS 587	ELEMENTARY ARABIC II
ARTS 547	COMMUNICATIONS INTERNSHIP

ARTS 555	BECOMING HUMAN
ARTS 551	AESTHETICS
ARTS 543	ART AND RELIGION

The defined concentration areas are as follows: Health Studies, Political Studies, Literature Studies, and Communication Studies.

Note that after this evaluation, the course ART AND RELIGION is up for debate about its status as an offering as an elective course. When we found data-based evidence to reduce the number of offered electives at the beginning of this section, ART AND RELIGION was selected to be retained. It has the lowest popularity ranking of the retained courses and only appeared once in the summary of Task 3's association rules. Additionally, the topic of art and religion do not fit well within our defined concentration areas.

Section 6

Summary of Experiments

The main goal of this mining experiment is to assist in the redesign of the Bachelor of Fine Arts (BFA) program at the Department of Arts of the University of Diploma Printing in Romanigstan due to stagnant enrollment. Possible reasons for stagnant enrollment in the BFA program include shifting disciplinary approaches and changing student interests and demographics. The intention of the redesign is to attract interest and boost enrollment numbers in the program.

The curriculum's courses are broken down into three areas: foundation, core, and elective. The curriculum redesign in this experiment concentrated on the elective courses. Unlike the foundation and core areas, the elective area does not have predetermined courses; the students are required to choose six out of the 33 elective course offerings at will.

During the discovery phase (Section 2), the two raw datasets are described and explored. The Curriculum dataset gave course groups in the foundation, core, and elective areas. The Student Enrollment dataset gave the historical enrollment data in the program. Seven problems with the raw datasets are found and preliminary solutions are given.

During the preprocessing stage (Section 3), the seven problems in the datasets are addressed. Using KNIME Analytics software, KNIME Workflows are created to fix the problems. The result of the preprocessing tasks was consistent and clean data stored in a number of datasets subsetted from the main two datasets. For instance, the Enrollment dataset was subsetted into Enrollment of Core courses, Enrollment of Foundation courses, and Enrollment of Elective courses. Using consistent and clean data is vital for the next stage of the experiment.

During the mining stage (Section 4), the two specific goals of the mining experiment are individually addressed and mining solutions are conducted in KNIME. To elicit information about reducing the number of offered electives, Task 1 found statistics about the Electives dataset and divided the courses into most popular and least popular. Task 2 used association rules for classification mining technique to find associations among the most popular elective courses; by using this technique, there was stronger evidence for which courses should be retained and which should be removed from offering. To elicit information about concentration areas, Task 3 used the association rule mining technique to find frequent patterns in the data. Because the students previously chose from a list of elective courses at will, information about elective courses frequently taken together is useful to determine concentration areas and groups of electives in each area.

The evaluation stage (Section 5) combined all of the results from the mining tasks. The data-based results turned into elicited information when combined with human logic and intuition. Discussions were created surrounding the information and data-based solutions to the pursued goals were found. Implementing the suggested solutions hopes to attract interest and boost enrollment numbers in the Bachelor of Fine Arts (BFA) program.

The following table contains the retained elective course offerings within their concentration area.

HEALTH STUDIES	POLITICAL STUDIES	LITERATURE STUDIES	COMMUNICATION STUDIES
AMERICAN HEALTH POLICY	CONTEMPORARY POLITICAL THOUGHT	19TH CENTURY BRITISH LITERATURE	ELEMENTARY ARABIC II
CELLULAR BIOLOGY AND BIOCHEMISTRY	BUSINESS GERMAN: A MICRO PERSPECTIVE	20TH CENTURY RUSSIAN LITERATURE: FICTION AND REALITY	COMMUNICATIONS INTERNSHIP
FRESHWATER ECOLOGY	COMPARATIVE POLITICS	AFRICAN AMERICAN LITERATURE	BECOMING HUMAN
ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL	FRANCE AND THE EUROPEAN UNION	DEVIL'S PACT LITERATURE/FILM	AESTHETICS
-	EUROPE IN A WIDER WORLD	BRITISH POETRY 1660-1914	ART AND RELIGION

Last Concerns & Further Suggestions

This experiment performed Tasks 1, 2, and 3 in order and then evaluated their individual results. In the future, if the mining experiment is replicated, I would suggest to perform Task 1 and Task 2, evaluate the results, and *then* perform Task 3. In this way, the elective courses removed from the course offering will not interfere with and will improve the results of Task 3.

As discussed at the end of Section 5, the course ART AND RELIGION is still up for debate about its status as an offered elective course within the Communication Studies concentration area. Further decision-making or data collection is required to make a final decision on the course. Suggested mining includes the association rules for classification technique with ART AND RELIGION as the Consequent to see which courses are strongly associated with it. Results from a preliminary mining task is included in **Appendix 4**.

Furthermore, the main goal of this mining experiment did not include altering existing courses or adding additional courses. However, it makes logical sense to also offer an ELEMENTARY ARABIC I course to take before ELEMENTARY ARABIC II. Or, combine them into simply ELEMENTARY ARABIC.

Appendix

Appendix 1
Top 20 and Bottom 20 Occurrences in the Raw Dataset

Student name	semester new	coursename
No. missings: 0	**No. missings:** 0	**No. missings:** 1
Top 20: Ed McMahon : 52 Harvey Golub : 52 Clive Barker : 48 Dave Barry : 44 Frances Lear : 42 Jimmy Buffett : 42 Philip Knight : 40 Leeza Gibbons : 39 Bonnie Blair : 36 Jodi Picoult : 36 Ed Asner : 33 Elizabeth Montgomery : 32 Judy Blume : 32 Don Mclean : 32 Laurie Metcalf : 30 Stephen King : 30 Daniel J. Boorstein : 30 Susan Elizabeth Phillips : 30 Debbie Reynolds : 28 Senator Dick Gephardt : 28	**Top 20:** Spring 2002 : 486 Fall 2004 : 417 Spring 2005 : 412 Spring 2001 : 398 Spring 2004 : 395 Fall 2003 : 349 Summer 2001 : 291 Summer 2004 : 285 Summer 2002 : 285 Fall 2005 : 271 Fall 2001 : 270 Spring 2003 : 248 Summer 2000 : 242 Fall 2002 : 202 Fall 2000 : 198 Summer 2003 : 151	**Top 20:** COMPUT LINEAR ALGEBRA : 411 Environmental Case Studies : 402 A WORLD AT WAR : 368 BEHAVIORAL PHARMACOLOGY : 354 ANALYTICAL MECHANICS : 350 CONTEMPORARY AFRICAN ART : 289 AMERICAN FOREIGN POLICY : 235 FOOD/FEAST ARCH OF TABLE : 151 AMERICAN HEALT POLICY : 127 EXPERIMENTAL WRITING SEM: The Ecology of Poetry : 125 ELEMENTARY FRENCH I : 102 EXPERIMENTAL WRITING SEM : 95 ELEMENTARY ARABIC II : 80 BEING HUMAN: Being Human: Biology, Culture & Human Diversity : 75 COMMUNICATIONS INTERNSHP : 67 CONTEMPORARY POL.THOUGHT : 61 19TH-CENT BRITISH LIT : 60 AMERICAN HEALTH POLICY : 55 FRESHWATER ECOLOGY : 55 CREATIVE WRITING: CREAT. WRITE: FICT/POET : 50
Bottom 20: RSally Jesse Raphael : 1 FNanette Fabray : 1 Bob Barker : 1 Tony Curtis : 1 Olympia Dukakis : 1 Bill Gates : 1 Haley Joel Osment : 1 Joan Rivers : 1 Jane Alexander : 1 Alistair Cooke : 1 Debbie Allen : 1 Diane Sawyer : 1 Eric Margolis : 1 Pat Schroeder : 1 Thomas H. Andrews : 1 Angus King : 1 Alan Alda : 1 Elmore Leonard : 1 David Bowie : 1 Robert Cormier : 1	**Bottom 20:**	**Bottom 20:** CREATIVE WRITING: CREAT./POETRY & NON-FICT : 1 COMMUNICATIONS INTERNSHiP : 1 ELEM BIBLICAL HEBREW II : 1 AMERICAN SOCIETY : 1 AFRICAN LANG. & CULTURE : 1 CONTEMPORARY AFRICAN-ART : 1 COMM, CHILDHOOD, & PLAY : 1 EARLY MESOPOTAM HIST/SOC : 1 ELEM MODERN HEBREW II : 1 EURO INT'L REL SINCE WW1 : 1 ACCEL INTERMEDIATE SPAN : 1 ACCEL INTERMD PORTUGUESE : 1 EURO INT'L REL SINCE WWI : 1 Community Based Research on Health Disparities : 1 CONVERSATION AND COMPOSI : 1 AMER REVOLUTION : 1 EURO INT'L REL SINCE WW One : 1 1000 YRS MUSICAL LISTENG: 1000 YRS MUSICAL LISTENG : 1 ART AND BUSINESS OF FILM : 1 CREAT.NON-FICTION WRIT: PEER TUTORING : 1

Appendix 2
Spelling alterations to the BFA Curriculum dataset

BFA CURRICULUM DATASET	
RAW	**ALTERED**
EXPERIMENTAL WRITING SEM: THE ECOLOGY OF POETRY	EXPERIMENTAL WRITING SEMINAR: THE ECOLOGY OF POETRY
DEVIL'S PACT LIT/FILM	DEVIL'S PACT LITERATURE/FILM
CONTEMPORARY POL. THOUGHT	CONTEMPORARY POLITICAL THOUGHT
AFRICAN-AMERICAN LIT: AFRICAN-AMER LIT:CHANGE	AFRICAN AMERICAN LITERATURE
COMM AND THE PRESIDENCY	COMMUNICATIONS AND THE PRESIDENCY
FRENCH THOUGHT TILL 1945	FRENCH THOUGHT BEFORE 1945
CONTEMP ART – 1945 TO PRESENT	CONTEMPORARY ART: 1945 TO PRESENT
EVIDENCE BASED CRIME AND JUSTICE POLICY	EVIDENCE-BASED CRIME AND JUSTICE POLICY

19-TH CENTURY BRITISH LITERATURE	19TH CENTURY BRITISH LITERATURE
AUGUSTAN CULTRAL REVOLUTION	AUGUSTAN CULTURAL REVOLUTION
ENVIRONMENTAL STUDIES RESEARCH SEMINAR JUNIOR LEVEL	ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL
ELEMENTARY GERMAN 1	ELEMENTARY GERMAN I
CELL. BIOL. & BIOCHEM.	CELLULAR BIOLOGY AND BIOCHEMISTRY
FRANCE & THE EUROP.UNION	FRANCE AND THE EUROPEAN UNION
ANALYZING THE POL WORLD	ANALYZING THE POLITICAL WORLD
EARLY MESOPOTAM HISTORY/SOCIETY	EARLY MESOPOTAMIAN HISTORY/SOCIETY
EARLY BALCAN HIST/SOC	EARLY BALKAN HISTORY/SOCIETY

Appendix 3
Spelling alterations to the Enrollment dataset

BFA CURRICULUM DATASET ALTERED SPELLING	ENROLLMENT DATASET MISSPELLING	# INSTANCES ALTERED
19TH CENTURY BRITISH LITERATURE	19TH-CENT BRITISH LIT	60
AFRICAN AMERICAN LITERATURE	AFRICAN-AMERICAN LIT AFRICAN-AMERICAN LIT: AFRICAN-AMERICAN LIT:CHANGE	1 20
AMERICAN HEALTH POLICY	AMERICAN HEALT POLICY	127
ANALYZING THE POLITICAL WORLD	ANALYZING THE POL WORLD	10
ART AND RELIGION	ART & RELIGION	8
ART: ANCIENT TO 1945	ART – ANCIENT TO 1945 ART – FROM ANCIENT TO 1945 ART ANCIENT TO 1945	9 25 8
AUGUSTAN CULTURAL REVOLUTION	AUGUSTAN CULTRL REVOL	13
BUSINESS GERMAN: A MICRO PERSPECTIVE	BUSINESS GERMAN – MICRO PERSPECTIVE BUSINESS GERMAN A MICRO PERSCPECTIVE BUSINESS GERMAN, A MICRO PERSPECTIVE	8 11 6
CELLULAR BIOLOGY AND BIOCHEMISTRY	CELL. BIOL. & BIOCHEM. CEL AND BIO AND BIOCHEMISTRY CEL BIO BIOCHEMISTRY CELL AND BIO AND BIOCHEMISTRY CELL BIOLOGY AND BIOCHEM CELL BIOL & BIOCHEM CELL BIOLOGY AND BIOCHEMISTRY CELL BIOLOGY & BIOCHEM CELL. BIOL. AND BIOCHEM.	14 2 21 6 23 13 15 9 6
COMMUNICATIONS AND THE PRESIDENCY	COMM AND THE PRESIDENCY COMM & THE PRESIDENCY	5 3
COMMUNICATIONS INTERNSHIP	COMMUNICATIONS INTERNSHP	67
COMPUTER LINEAR ALGEBRA	COMPUT LINEAR ALGEBRA	411
CONTEMPORARY ART: 1945 TO PRESENT	CONTEMP ART – 1945 TO PRESENT CONTEMP ART:1945 TO PRESENT	4 4
CONTEMPORARY AFRICAN ART	CONTEMPORARY AFRICAN-ART	1
CONTEMPORARY POLITICAL THOUGHT	CONTEMPORARY POL.THOUGHT	61
CONTEMPORARY SOCIO THEORY	CONTEMPORARY SOCIO THEOR	1
DEVIL'S PACT LITERATURE/FILM	DEVIL'S PACT LIT/FILM	16

EARLY MESOPOTAMIAN HISTORY/SOCIETY	EARLY MESOPOTAM HIST – SOC EARLY MESOPOTAM HIST/SOC EARLY MESOPOTAMIAN HIST - SOC	1 1 9
ELEMENTARY GERMAN I	ELEMENTARY GERMAN 1	2
ENVIRONMENTAL STUDIES RESEARCH SEMINAR, JUNIOR LEVEL	ENVIRONMENTAL STUDIES RESEARCH SEMINAR JUNIOR LEVEL ENVIRONMENTAL STUDIES RESEARCH SEMINAR FOR JUNIORS	16 5
EVIDENCE-BASED CRIME AND JUSTICE POLICY	EVIDENCED BASED CRIME & JUSTICE POLICY EVIDENCED BASED CRIME AND JUSTICE POLICY	14 2
EXPERIMENTAL WRITING SEMINAR: THE ECOLOGY OF POETRY	EXPERIMENTAL WRITING SEM: THE ECOLOGY OF POETRY EXPERIMENTAL WRITING SEM	125 95
FRANCE AND THE EUROPEAN UNION	FRANCE & THE EUROP.UNION	23

Appendix 4

Preliminary mining to determine if ART AND RELIGION should be retained in the elective course offerings.

Mining parameters

Minimum set size: 2
Minimum support: 5%
Minimum rule confidence: 5

Association Rules of Consequent ART AND RELIGION by most actionable (highest Lift).

Antecedent	Consequent	Support Count	Confidence %	Lift
COMPARATIVE POLITICS	ART AND RELIGION	2	8.7	2.539
ELEMENTARY ARABIC II	ART AND RELIGION	5	7.81	2.281
COMMUNICATIONS INTERNSHIP	ART AND RELIGION	3	6.67	1.947
FRANCE AND THE EUROPEAN UNION	ART AND RELIGION	1	6.25	1.825
19TH CENTURY BRITISH LITERATURE	ART AND RELIGION	2	5.13	1.497