ISSN: 2473-4901
v12 n6847

# Teaching Business Students to Use Data Mining Methods for Decision Support

Liuliu Fu
Lfu8@calstatela.edu
California State University
Los Angeles, CA 90032, USA

Lusi Li
lli57@calstatela.edu
California State University Los Angeles
Los Angeles, CA 90032

Jinzhi Li
jli037@odu.edu
Old Dominion University
Norfolk, VA 23529, USA

## Abstract

This paper demonstrates how to adopt Herbert Simon's three-step decision-making process: intelligence (understand the problem), design (build analysis models), and choice (make decisions), to guide students through a complete data analysis report. The student project is built on real datasets from the author's earlier research on a Faceted Multimedia Classification System. The paper illustrates how to help students apply Simon's framework to analyze user experience data and develop evidence-based recommendations for system improvement. The project is intended for Information Systems students who do not have advanced programming or statistical skills. It focuses on applying decision logic and procedures using tools the students have already learned. The case connects data analysis with managerial decision making, helping students develop the skills to support technology evaluation in practice.

**Keywords:** Business intelligence and analytics, Decision Support Systems, Data mining, Quantitative research, Qualitative research, Survey research

ISSN: 2473-4901
v12 n6847

# Teaching Business Students to Use Data Mining Methods for Decision Support

*Liuliu Fu, Lusi Li and Jinzhi Li*

## 1. INTRODUCTION

This paper demonstrates how data can be analyzed to support decisions about system evaluation and improvement. The teaching case draws on datasets collected during a research project that studied the implementation of a Faceted Multimedia Classification System (Fu et al., 2010, 2011). The three datasets include survey responses from approximately 100 students, gathered before and after the experiment, together with logs of how those students clicked and interacted with the Faceted system while the experiment was running.

We now describe the Faceted system in more detail and explain how the three datasets were collected. The goal of the Faceted system was to help users organize, tag, and retrieve information more effectively while supporting interaction and collaboration among users. In the underlying study, students completed a series of tasks in the system while their interactions and actions were logged (Fu et al., 2011). To capture user perceptions, two surveys were administered: one before and one after the experiment. The first measured students' initial understanding of social networking and tagging technologies, and the second recorded their opinions after they had used the system, a faceted interface for the African American History Image Collection shown in Figure 1 (Fu et al., 2010). Each student was assigned 20 image items and asked to complete tasks such as adding tags, classifying the images into appropriate categories across different facets, and voting on classifications created by other students. Together, the pre-survey responses, post-survey responses, and system activity logs form a complete dataset that supports analysis from a Decision Support System (DSS) perspective.


**Figure 1. Screenshot of the Main Page of Faceted System (2010)**

The primary objective of this teaching case is to help students apply DSS methods to generate descriptive, predictive, and diagnostic insights that support managerial and design decisions in business analysis contexts, using the Faceted Multimedia Classification System and its datasets as the working example. Students follow Simon's decision-making process to carry out the analysis, and the project requires them to propose and answer DSS-oriented questions such as:

- From the data, how well did students understand social networking systems?
- How well is the new system designed, and what are its weaknesses?

 ISSN: 2473-4901
v12 n6847

- How should missing or erroneous data, such as an invalid student ID, be handled?
- Are there patterns or connections among the different student activities?
- What can be inferred about the system from analyzing the students' activities?
- Is there alignment among the three datasets, and how can the three data tables be linked?

Students are free to select from a range of decision support tools, such as Excel, SPSS, SAS, or SAP Analytics Cloud. No particular tool or software package is required; students may use whatever they have learned in other classes, depending on their background and experience.

Different analytical methods, such as descriptive statistics, clustering, and classification, can be applied to turn raw data into knowledge. Students should treat the project as a decision-making process that includes data cleaning, transformation, model selection, and interpretation of findings. Because the course serves undergraduate business students in Information Systems, who may not have the programming skills or advanced statistical training typical of engineering majors, the emphasis is on the logic and sequence of the decision-making process rather than on complex data mining techniques or programming.

All procedures and results should be documented clearly and supported with tables, charts, or figures. The final report should be 10–15 pages, written collaboratively, and should identify each group member's specific responsibilities. Students submit both the cleaned and recoded datasets (e.g., the pre-experiment survey file, survey_pre, and the system activity log file, user_activity) and the written report. Proper references and visual evidence are expected to support the analyses.

Through this case, students gain practical experience in using DSS methods to transform raw data into insights that inform design evaluation and system improvement. The exercise helps bridge the gap between data analysis and managerial decision making in technology-supported learning environments.

## 2. THE PROCESS OF DECISION MAKING

Herbert A. Simon received the Nobel Prize in Economic Sciences in 1978 and is best known for his work on decision-making processes within organizations (Augier & March, 2004). He proposed a widely used model that treats decision processes as structured, iterative activities rather than single events. The model is especially relevant to Decision Support Systems, where data and analytical tools support each phase of decision making.

Simon described decision making as a process with three main phases: intelligence, design, and choice (Simon, 1966), as illustrated in Figure 2. The intelligence phase focuses on identifying and understanding problems or opportunities; the design phase develops and analyzes possible models or courses of action; and the choice phase selects a preferred alternative based on those models.

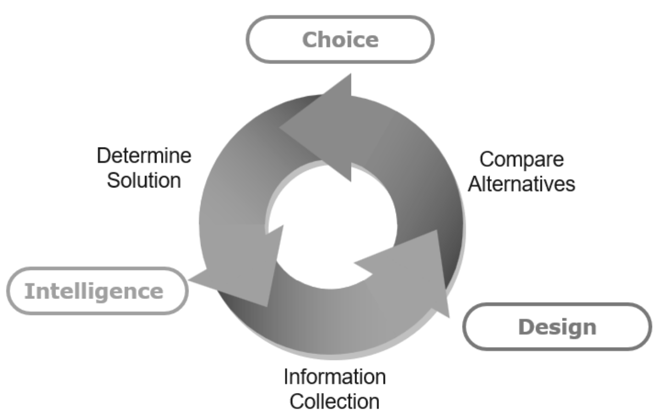


**Figure 2. Simon's Decision-Making Process**

ISSN: 2473-4901
v12 n6847

Huber and McDaniel (1986) extended Simon's model by adding two further phases: implementation and monitoring. In their version, the implementation phase turns the decision reached in the choice phase into feasible actions and organizational practice, while the monitoring phase covers follow-up evaluation, including measuring outcomes, gathering performance feedback, and making adjustments that inform and shape later decision-making cycles.

In this teaching case, however, students do not carry out real-world implementation or monitoring of the Faceted Multimedia Classification System themselves. Their task is instead to develop practical implementation plans based on their analytical recommendations, such as interface redesign ideas, training protocols, or phased rollouts, which are optional. The case therefore adopts Simon's classic 1966 framework, which emphasizes the core intelligence, design, and choice phases as the foundation for decision support analysis, and treats implementation and monitoring conceptually rather than operationally. Although DSS is the primary topic, the case can also be used in other courses where students learn to transform data into managerial or design recommendations.

### 2.1 The Intelligence Phase

In this teaching case, the intelligence phase begins with a fundamental managerial question: does the proposed Faceted Multimedia Classification System genuinely improve students' learning, information-tagging behavior, and social networking interactions within the course environment? The phase requires students to identify potential problems and opportunities systematically: whether students used the system as intended, whether they found it helpful, whether its functions aligned with instructional goals, and whether prior technical experience with social platforms was associated with how students used the system and how they evaluated it.

The underlying research on the Faceted system produced three datasets: pre-experiment survey responses capturing students' initial expectations and prior experience with similar technologies, post-experiment survey responses reflecting their satisfaction and perceived usefulness after using the system, and detailed activity logs documenting how students worked through the tagging and classification tasks. Together, these sources allow students to identify issues such as usability barriers, patterns of low or uneven engagement across student groups, and mismatches between what students expected the system to do and what they experienced.

In the intelligence phase, students work in groups to become familiar with the datasets, clarifying the meaning of each variable and making sure they understand what each field represents in terms of student experience, system evaluation, and behavior. They also review the underlying research that produced the data, including who was surveyed, when the measurements were taken, and how the activity logs were generated during system use. On this basis, they begin to identify the main research questions to examine, such as which aspects of system usage or student perception should be the focus of later analysis and decision making.

### 2.2 The Design Phase

In the design phase, students shift from understanding the Faceted Multimedia Classification System and how its data were collected to selecting and constructing appropriate models. Beyond asking what happened while users worked in the system and answered the survey questions, students begin to consider how to represent the problem in a structured way, what assumptions to make, and which variables matter most for decision making. This means moving from broad questions to precise analytical plans that can be tested and compared.

A key part of the design phase is deciding how to preprocess the data so that it is suitable for analysis. This may include handling missing or erroneous student IDs between the pre-survey and post-survey, merging or splitting categories, recoding survey items into meaningful scales by converting text responses into numeric values, and integrating the three data tables (pre-survey, post-survey, and activity logs) into a coherent dataset. Students also choose appropriate analytical methods: descriptive statistics to summarize usage and perceptions, clustering to discover user groups, correlation analysis to examine how strongly and in what direction pairs of variables are related, or classification methods to predict outcomes from user characteristics and behaviors.

ISSN: 2473-4901
v12 n6847

By specifying and combining these approaches, comparing user groups, relating perceptions to behaviors, and identifying patterns in the activity logs, students are in effect designing alternative decision models. Each model offers a different view of the data and can lead to different recommendations about the system's design, adoption, and deployment. In this way, the design phase lays out a structured set of options that are evaluated and selected in the choice phase.

### 2.3 The Choice Phase

In the choice phase, students move from analysis to action. Having produced several feasible models, they compare the results, weigh the trade-offs, and decide which course of action best addresses the problems and objectives they defined.

During the semester, students enter the choice phase once they have produced and interpreted their key analytical results, such as clusters of different user types, relationships between the perceived usefulness of functions and user evaluations, or patterns of errors. They use these findings to compare concrete decision options, for example: "adopt the system with minor adjustments," "redesign or simplify specific features," "provide targeted training for certain user groups," or "limit use of the system to particular courses or contexts." At this point, students must argue not only what should be done, but also why one option is more suitable than the others, based on the evidence they have generated.

Students complete the choice phase when they formulate clear, prioritized recommendations for the Faceted Multimedia Classification System and support each recommendation with specific results from the pre- and post-survey data and the activity logs. This means pointing to particular patterns. Is engagement with some functions unusually low or high? Are there strong links between users' prior experience and their evaluations of the system? Do certain functions generate frequent requests for help? Students then explain how these patterns justify their preferred option. By presenting recommendations clearly and tying them to the data, students show that their decisions are logical and feasible.

## 3. CASES OF STUDENT PROJECTS

In this section, we summarize selected examples from student projects completed in the Fall 2025 semester. They show how the decision-making framework was used to analyze and evaluate the Faceted Multimedia Classification System.

We illustrate the process by working through each of the three phases of decision making, intelligence, design, and choice, and showing what students did in each. For the intelligence phase, we describe how students worked to understand the problem and the data. For the design phase, we show how they selected models and ran different analyses. For the choice phase, we explain how they used their results to draw conclusions and make recommendations about the system.

Organizing the student work in this step-by-step way makes clear not only which analyses were used, but also how those analyses helped students reach better decisions about how the system should be evaluated, improved, or used in the future.

### 3.1 The Intelligence Phase

In the intelligence phase, students worked to understand both the data and the background of the Faceted Multimedia Classification System. They first examined the three datasets (pre-survey, post-survey, and action logs) and clarified what each column meant, how each piece of data had been collected, and how the three files could be connected for each student. This included understanding who answered the surveys, when the surveys were taken, and what each type of action in the log, such as clicking, tagging, or opening a page, represented in the system.

To support this work, screenshots of the Faceted Multimedia Classification System (Fu et al., 2011) were shown in class to demonstrate how users' actions were captured and turned into log data (Figure 3). Students also learned how to code and recode key concepts, such as familiarity, usage level, ease of use, and satisfaction, into simple numerical scales that could be analyzed later. On this basis, they moved from the broad question "Is the system effective?" to more specific, data-based questions that

 ISSN: 2473-4901
v12 n6847

guided the rest of their analysis.


**Figure 3. Screenshot of Classification Page of Faceted System**

Student progress reports showed clearly how the intelligence phase worked in a real classroom setting. One group stated that its goal was to use all three datasets to "better understand how students interacted with the facet classification system." The group turned this broad aim into specific research questions to explore first, such as:

- How comfortable did students feel with technology and tagging before and after using the system?
- What patterns appeared in the way they used the system?
- Which specific features worked well, and which did not?
- How were students' expectations connected to their behavior and final satisfaction?

Another group began by describing the setting: college students using a tagging and classification tool. The group explained what each dataset showed, the pre-survey for starting familiarity, the post-survey for ease of use after the exercise, and the action logs for engagement measured through activity counts. It then listed questions about initial familiarity levels, how easy the system felt, engagement patterns, connections between the datasets, and how well or poorly the system appeared to be designed.

Some groups went further. One documented step by step how it combined the three files into a single ready-to-analyze table. The group described how survey answers were coded into numbers on the pre- and post-scales, reported baseline statistics on familiarity and usage, and organized its whole report around simple guiding questions: what students knew before starting, what patterns appeared in their actions, and how well the system was designed from the students' point of view.

In these examples, students completed the intelligence phase by studying the data carefully, reconstructing how the original research had been conducted, and formulating clear research questions. Those questions then guided their work in the design and choice phases.

### 3.2 The Design Phase

In this section, the design phase turns the initial questions into concrete data preparation steps, analytical models, and visual tools that support decision making about the Faceted Multimedia Classification System. Because the material is extensive, one submitted report is used throughout as a worked example of how the design phase was carried out.

#### 3.2.1 Data preparation and integration

The group first cleaned and recoded the pre-survey, post-survey, and action-log datasets so that they could be analyzed together. Text familiarity responses (e.g., "Very familiar," "Never heard of this") were

ISSN: 2473-4901
v12 n6847

converted into numeric scores and summed into a composite Tech Score, which was then split into Low, Medium, and High Tech groups using percentile cut-offs and Excel formulas. Pre- and post-survey totals, originally on different scales (1–4 across 23 items versus 1–5 across 12 items), were normalized to 0–1 scores (PreNormScore and PostNormScore), and a ScoreChange variable was created for change analysis. This group wrote Python code to recode the pre-survey, whereas other groups relied on Excel formulas or even recoded responses manually.

Student-level tables were then linked and merged on StudentID so that each row combined tech familiarity, activity counts (including Total_Activities and per-action counts), and post-survey ratings in a single analysis-ready file, supported by a data dictionary.

### 3.2.2 Descriptive grouping by Tech Level

Students were divided into three Tech Level groups (Low, Medium, and High) based on the composite Tech Score, so that behavior and help-file usage could be compared across levels of familiarity.

Baseline descriptive statistics, such as average familiarity, frequency of use, the share of students who tagged content frequently, and prior Blackboard experience, were computed to characterize who the students were before they used the system.

### 3.2.3 Exploratory and diagnostic analyses

For RQ1 (How does pre-survey technical familiarity relate to student behavior in the system?), the group linked Tech Level to the action logs, created visualizations of average pre-survey scores, total actions, and help-file usage by Tech Level, and interpreted these patterns to diagnose the confidence, pace, and support needs of different user groups. Figure 4 shows the total number of system actions performed by each individual student, grouped by Tech Level. Because the chart displays the distribution of activity counts across individual students rather than a single value per group, it allowed the group to examine the range and variability of engagement instead of assuming a single group average. Engagement varied considerably both within and across the three groups, with some High Tech and some Low Tech students recording especially high activity counts.

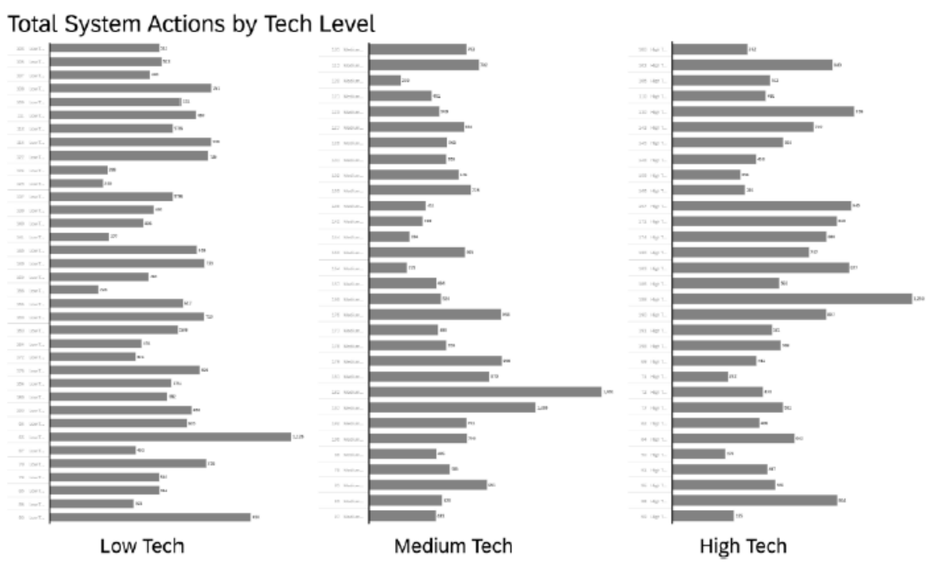


**Figure 4. Total System Actions by Tech Level**

For RQ2 (What system features show signs of usability issues?), the group worked only from the Actions-by-User dataset. It first built a pivot table of StudentID by ActionType, filled blank cells with zeros, and computed a correlation matrix, focusing on correlations with "view help file." A heatmap created through conditional formatting highlighted the actions most closely associated with help-file clicks and revealed navigation and category-structure tasks as the problem areas. The group then identified the "Help Users"

 ISSN: 2473-4901
v12 n6847

who had clicked help at least once, built a pivot table on that subset alone, and ranked actions by frequency to see which tasks confused help-seeking students most. Figure 5 reports the strongest correlations between system activities and help-file usage; actions related to navigating the category structure show the highest values.

| Activities | Correlation with View Help File |
|---|---|
| rate association | 0.256 |
| view all categories of facet | 0.208 |
| expand category | 0.173 |
| rate item | 0.154 |
| collapse category | 0.149 |
| go to page | 0.111 |
| remove association | 0.110 |
| open classification page | 0.108 |
| refine category/facet | 0.092 |
| view description | 0.094 |
| rename category | 0.087 |
| search terms | 0.014 |

**Figure 5. Correlation between System Activities and Help File Usage**

### 3.2.4 Predictive modeling

For RQ3 (To what extent can students' pre-survey expectations and system activity levels predict their post-survey satisfaction?), the group selected simple and multiple regression models to test whether expectations and engagement predict satisfaction.

The group began by calculating Activity Volume, the total number of actions per student, from the action logs. It then ran three regressions: PostNormScore on PreNormScore alone; PostNormScore on Activity Volume alone; and a multiple regression using both PreNormScore and Activity Volume as predictors. Finally, it interpreted the $R^2$ values and coefficients to evaluate predictive strength. Expectations and activity turned out to have limited predictive power (low $R^2$ values), suggesting that unmeasured design factors such as interface clarity or task instructions influenced satisfaction more strongly than the variables captured in these models.

### 3.2.5 Change and comparison analysis

For RQ4 (How do pre-survey expectations compare to post-survey satisfaction?), the group examined how students' attitudes changed from before to after using the system. It first created a ScoreChange variable by subtracting PreNormScore from PostNormScore, which showed whether students felt better, worse, or the same after the activity.

The group then produced two charts: a scatterplot of PreNormScore against PostNormScore, to see how starting expectations lined up with final satisfaction and to spot any movement or patterns; and a bar chart of the 15 largest positive ScoreChange values, to highlight the students who improved the most.

These charts showed that many students ended up more satisfied than they had expected at the outset, especially those who began with average, middle-range familiarity. This suggests that the system built confidence even among students who were unsure of it at first.

### 3.2.6 Tools and implementation details

In the design phase, the group used several tools for data preparation and analysis: Excel (pivot tables, percentiles, correlations, regressions, and conditional formatting), SAP Analytics Cloud, and a limited amount of Python and pandas.

 

 ISSN: 2473-4901
v12 n6847

Taken together, these steps show how the design phase works in practice. Students start with broad research questions, such as "Does prior familiarity affect system use?" or "Which features confuse users?" They then break these questions down into practical actions. They transformed the raw data by cleaning it, recoding text answers into numbers, normalizing survey scales that differed, and creating new variables such as Tech Score, Activity Volume, and ScoreChange. They grouped students into Low, Medium, and High Tech categories so that behavior could be compared across types of users. They built correlation matrices and heatmaps to discover which system actions were linked to confusion, such as clicking the help file. And they created simple and multiple regression models to test predictions about satisfaction.

Each of these choices produces a different model for understanding the tagging system. One model might show that Low Tech students need more help, another that navigation features cause confusion, and a third that expectations predict satisfaction better than activity levels do. By building this set of analytical models, students generated several evidence-based explanations of where the system works well, where it does not, and which specific improvements are needed.

### 3.3 The Choice Phase

In the choice phase, the student groups used their analytical results to reach and justify concrete conclusions about the Faceted Multimedia Classification System. The group whose design phase we followed above concluded that the system was worth keeping for learning purposes but needed targeted redesign. It found that higher-tech students used the system more efficiently and needed less help, that navigation and category-structure features were not intuitive and should be redesigned, and that many students ended up more satisfied than they had expected despite these issues, which confirmed a positive educational impact. The group therefore recommended retaining the system while improving navigation, clarifying the classification structure, and adding better guidance, such as tooltips and clearer instructions, for less experienced users.

Another group concluded that the exercise successfully helped students work together and understand the basics of social systems, but it also identified design problems. Its results showed that users handled the basic features easily yet had trouble with more advanced ones: the system was simple to use overall, but problems appeared in tasks involving creativity, such as making new items, and reliability, such as working without errors. The group also noticed patterns linking navigation actions to rating actions, along with K-means clusters that grouped different types of users. These findings led it to recommend modest design changes that would better support different kinds of users and make the advanced features easier to use.

Other groups reported further findings. One report noted that most students behaved like browsers, simply looking around, rather than builders who create tags, associations, or categories. It concluded that prior experience helped only slightly with productive actions, and that students who used the help feature more often felt more stuck, which suggests that the timing of help and the way it is delivered matter more than the clarity of the help text alone. The group therefore recommended two improvements: first, to encourage more productive actions through better guidance and support for tags, associations, and categories; and second, to redesign the help feature so that it appears at the right moment and works with the interface, rather than simply adding more text.

These examples show how the choice phase works in real student projects. Each group examined its results carefully, decided whether the system should be kept as it was, changed, or supplemented with additional support, and then proposed specific changes. Every recommendation came directly from evidence in the data, whether redesigning confusing parts of the interface, improving guidance for different types of students, or strengthening learning support.

## 4. CONCLUSION

This case shows how to guide business students through Simon's decision-making process as they analyze data from a faceted tagging and classification system and make practical suggestions for improvement. The process has three phases: intelligence (understanding the problem and the data), design (creating analysis models), and choice (deciding what to do based on the results). Students followed these phases successfully despite having no expertise in programming or advanced statistics.

ISSN: 2473-4901
v12 n6847

Designed for Information Systems students, the case emphasizes the thinking process, its logic and sequence, rather than advanced technical skill. Students learned to turn raw user data into real decisions about technology design, and they gained hands-on experience with Decision Support Systems that connect data analysis to practical business choices.

The Fall 2025 examples show that the approach works in a real classroom. Students produced useful insights and recommendations using a variety of tools, demonstrating that business students can apply DSS methods to evaluate educational technology and propose improvements even without advanced training. The structured sequence of intelligence, design, and choice helps them think clearly and decide on the basis of evidence, a key skill for their future careers.

## 5. REFERENCES

Augier, M., & March, J. G. (Eds.). (2004). Models of a man: Essays in memory of Herbert A. Simon. *MIT Press.*

Fu, L., Maly, K., Wu, H., & Zubair, M. (2010). Building Dynamic Image Collections from Internet. In *DH* (pp. 156-158).

Fu, L., Maly, K., Rasnick, E., Wu, H., & Zubair, M. (2011). User Experiments of a Social, Faceted Multimedia Classification System.

Huber, G. P., & McDaniel, R. R. (1986). The decision-making paradigm of organizational design. *Management science*, 32(5), 572-589.

Simon, H. A. (1966). Theories of decision-making in economics and behavioural science. In *Surveys of economic theory: Resource allocation* (pp. 1-28). London: Palgrave Macmillan UK.