Introducing Data Mining Techniques and Software Engineering to High School Science Students

Bartley D. Richardson, Karen C. Davis, Michelle Daniel Beach
richardbe@ececs.uc.edu, karen.davis@uc.edu, mbeach@berea.k12.oh.us

Abstract – This paper describes the activities of a Computer Science doctoral student and a Secondary Education masters student in the design, development, and implementation of a lesson for a high school science class. The graduate students, called Fellows, worked in secondary classrooms in the Cincinnati Public School District as a part of Project STEP at the University of Cincinnati, which is funded by the National Science Foundation GK-12 Program. The Fellows formed partnerships with secondary math and science teachers to generate new lessons, activities, and resources to enhance the STEM skills of high school students. Additionally, the Fellows used their engineering expertise to bring authentic, inquiry-based learning experiences into the classroom and introduced engineering concepts to underserved student populations. This paper highlights a lesson that integrates data mining and software engineering into a physical science lesson that focuses on the periodic table. Included in the paper are techniques used by the Fellows, reactions and feedback from the students, and observations and reflections by the Fellows regarding aspects of the activity that had the most impact on student learning.

Index Terms – Data mining, STEM education, secondary science, NSF GK-12

INTRODUCTION

As the number and quality of trained engineers around the world increases, the United States faces increasing challenges to stay at the forefront of the global technology race. An important determinant of our success will be the quality and quantity of engineers we produce from our colleges and universities. In order to produce these quality engineers, recruitment of talented high school students to engineering and technical fields is a high priority. Unfortunately, secondary educators, those individuals with the most interaction with students, are often unable to provide the necessary information to interested students since they lack the background, time, and resources to effectively introduce engineering concepts and careers.

In order to address these issues in our secondary schools, Project STEP (Science and Technology Enhancement Program) was developed at the University of Cincinnati with funding from the National Science Foundation. There are three primary goals of Project STEP: (1) to produce scientists and engineers who are experienced in bringing their technical and educational expertise into the classroom, (2) to design, develop, and implement hands-on activities and inquiry-based lessons that are relevant to everyday life, and (3) to encourage secondary students to consider engineering as a future college major and an eventual profession [1].

The Fellows are the key component to Project STEP. Fellows are engineering and education graduate students who bring their research and teaching expertise into secondary classrooms through the use of lessons and hands-on activities. During these lessons, high school students are exposed to engineering concepts and careers while they learn in an authentic, inquiry-based method. Additionally, the lessons are aligned with State of Ohio and national standards, thus the students are learning standard material in new and exciting ways.

This paper provides details about instructional activities developed for secondary students by two Fellows, one with an Electrical and Computer Engineering background and the other with an education background. The lesson highlighted focuses on introducing data mining and software engineering concepts to physical science students. The arrangement of elements on the periodic table was likened to the organization of a grocery store. By using a concept students are familiar with, a frequent buyer card, the Fellows were successful in explaining how companies use data collected from customers to organize and price items. Using a grocery store layout as a starting point, the Fellows were able to explain how elements in the periodic table also have a specific arrangement and are close to other similar elements. This paper also presents student feedback and observations and reflections by the Fellows about the aspects of the lesson that had the most impact on student learning and interest in engineering.

RATIONALE

This lesson, titled “Mega Mining Mart,” was designed to either supplement or fully replace an initial discussion on the arrangement of elements on periodic table to physical science students in a high school or junior high setting. The goal is for students to identify trends and understand correlations in a familiar setting (a grocery story) and then to apply this understanding to trends and correlations on the periodic table. While discussing a grocery store, the idea of introducing data mining (through the use of frequent buyer cards) and software engineering is a natural extension to the grocery store analogy being presented. Students’ prior learning includes knowledge about elements, compounds,
states of matter, chemical changes, physical changes, and density.

“Mega Mining Mart” was designed to be a hands-on activity. Students were provided with a blank store layout consisting of empty shelves, cash register locations, and entrance and exit points. They were also given paper cutouts (manipulatives) of various sizes that represented 50 items typically found in a grocery store. These items included dry goods, refrigerated items, freezer items, fresh produce, and toiletries and other sundry items. In addition, they were provided six data mining associations. The associations were of the form “Customers who buy X also buy Y.” Students created categories for the items and then placed them in the store in a way that made sense to the customer (similar items grouped together) and satisfied the data mining associations. A more complete discussion of this is presented in subsequent sections.

While there are other GK-12 programs like Project STEP around the country, most of them do not have searchable repositories of lessons. The analysis of their websites and repositories are beyond the scope of this paper. A search was performed on various well-known K-12 repositories but resulted in no results similar to this lesson at the time of publication.

ENGINEERING COMPONENTS

I. Motivation

This lesson was designed as a replacement for an introductory lesson about trends in the periodic table. Prior to lesson development, the Fellows polled students about their experiences, routines, and family responsibilities. It was clear that a large majority of students were intimately familiar with the arrangement of their local grocery store. Many of them indicated that they were either the primary shopper for their family or frequently helped the primary shopper. Since most stores collect customer purchase information via a frequent buyer card, this was an opportunity to introduce students to the concept of data mining and the area of software engineering.

II. Data Mining Concept

Data mining is used by software engineers to find unique and previously unknown statistics within a set of collected data. This data is typically transaction-based (such as that from a grocery store) and is stable (does not change once it is collected). With more stores offering customers pricing incentives to use their frequent buyer card, it is easy for these stores to amass a large amount of transaction data. This data is then mined (searched) for associations. For example, if a store finds that customers typically buy a particular type of peanut butter with a certain type of jelly, they may be inclined to place the related peanut butter and jelly in close proximity to each other on the shelf. The hope is that more customers who see either item will then be inclined to also purchase the other.

Data mining can also be used to increase profits between seemingly unrelated items. By analyzing the transaction (purchasing) data of many customers, a store might discover that a large percentage of customers who purchase bananas also purchase bread and milk. The store then may drop prices on one item (bananas, for example) and slightly raise prices on the other two items. Regardless of the final action, the end result of data mining is always the same – discovery of previously unknown connections between seemingly unrelated items or instances of data. Even though a direct relation is not obvious, data mining surfaces these correlations.

The concept of data mining was a good extension to a discussion about the periodic table. Initially, students are unfamiliar with the connections and trends among elements. It is only through doing more investigation that these trends become apparent. The goal of this work was to introduce students to the concept of data mining and software engineering while still teaching them about the periodic table.

II. Software Engineering

In addition to teaching students about the periodic table and data mining, another goal was to acclimate them to a potential career that utilizes math and science. If at all possible, this career is closely tied to the concepts discussed during the lesson. For this lesson, a software engineering career was selected for two reasons. First, it is an engineering specialty that few students know about. One of our goals as Fellows was to give students as much information about engineering, math, and science careers as possible. Second, if a student were interested in the engineering concept (data mining) introduced, a career as a software engineer would allow them to further pursue these interests. The career sheet was one of many given to the students throughout the course of the academic year. Students were instructed to compile their career sheets into a binder where they could be easily referenced throughout the year and as they start planning for their college careers.

In addition to the career sheet, the Fellows discussed the concept of software engineering and how it directly relates to the computer science and engineering fields. An opportunity for students to interact with the Fellows was provided via a question-and-answer session. Later in the academic year, students were given the opportunity to interact with a professional engineer. This professional reinforced the concepts discussed by the Fellows and those given to the student on the software engineering career sheet.

LESSON DESIGN AND DEVELOPMENT

The lesson, titled “The Mega Mining Mart,” was developed for a 9th grade physical science class. It is an extension of an introductory lesson on the periodic table and element trends and was designed to fill two 70 minute class periods. The following information presents specifics about the specifics
of the lesson and the materials developed for its implementation.

I. Objectives

The objectives of this lesson were selected to meet state academic content standards [2]. Students will
- apply methods of inventive problem solving to plan the arrangement of a store using their personal knowledge of grocery stores and the data mining relationships provided to them,
- analyze trends in the arrangement of their store,
- relate the arrangement of the items in their store to the arrangement of elements in the periodic table,
- recognize that the periodic table was formed as a result of the repeating pattern of electron configurations and similar chemical and physical properties, and
- understand how data mining impacts society.

The Fellows working with a science teacher in the targeted high school selected these objectives. Since instruction time with the students is limited, any lesson that sought to introduce engineering concepts was also required to meet the necessary science standards so that the required content is still covered in the classroom.

II. State Standards

The “Mega Mining Mart” lesson was designed to meet multiple state standards in both science and technology that target the 9th grade. Below is a brief description of these standards.

Ohio’s Academic Content Standards in science are divided into six main groups – Earth and Space Sciences, Life Sciences, Physical Sciences, Science and Technology, Scientific Inquiry, and Scientific Ways of Knowing [2]. This lesson meets Benchmark A of the Physical Sciences standard and Benchmark B of the Science and Technology standard. Benchmark A requires students to describe that matter is made of minute particles called atoms, atoms are comprised of even smaller components, and explain the structure and properties of atoms. Benchmark B requires students to explain that science and technology are interdependent (each drives the other).

Ohio’s Academic Content Standards in technology are divided into seven main groups – Nature of Technology, Technology and Society Interaction, Technology for Productivity Applications, Technology and Communication Applications, Technology and Information Literacy, Design, and Designed World [3]. This lesson meets Benchmark A of Technology and Society Interaction, Benchmark A of Technology for Productivity Application, Benchmark A of Design, and Benchmark E of Designed World. These benchmarks require students to
- interpret and practice responsible citizenship relative to technology,
- integrate conceptual knowledge of technology systems in determining practical application for learning and problem solving,
- identify and produce a product or system using a design process, and
- classify, demonstrate, examine, and appraise information and communication technologies.

These benchmarks are measured through various State indicators. There are multiple indicators for each benchmark. The main points of these indicators require students to
- understand that the development of technology may be influenced by societal opinions and demands, in addition to corporate cultures,
- explain and apply the methods and tools of inventive problem-solving to develop and produce a product, and
- describe the careers available in information and communication technological systems and the training needed to pursue them.

While these standards are specific to the State of Ohio, the State of Ohio’s Academic Content Standards are aligned with national standards, so this lesson can be easily implemented anywhere in the United States.

III. Background Knowledge

Before completing this lesson, students should already have basic knowledge about elements and a brief understanding that elements are composed of neutrons, protons, and electrons. In addition, they should have basic knowledge about food items common to grocery stores and a typical grocery store layout. Prior to the start of the actual lesson, students will be given a description of data mining, a common example of its use, and a career that relates to the concepts they will be learning.

IV. Materials

The “Mega Mining Mart” lesson requires minimal materials. This allows the teacher to easily implement the lesson regardless of the instructional setting. The Fellows developed a packet that the students use throughout the lesson. This packet includes step-by-step instructions, all necessary worksheets, a career sheet, and a feedback survey. The career sheet describes software engineering, and it includes information about what software engineers do, their typical salary, job outlook through 2012, necessary education, what to study in high school, and state colleges and universities that offer software engineering courses. Throughout the year, students collect multiple career sheets and place them in a folder. They are then able to reference them whenever necessary and use them to make informed decisions about career paths.

In addition to the “Mega Mining Mart” packet, the Fellows also developed a store layout template and templates for typical grocery store items. All of the necessary data mining association rules are contained in the “Mega Mining Mart” packet, and all materials are available for download at the Project STEP website [1].
**LESSON ACTIVITIES**

The “Mega Mining Mart” was designed to be a group project. Students should be divided into groups of two or three (if resources are limited). The section below describes, in detail, the activities and flow of the lesson in the classroom. Observations and reflections by the Fellows on this implementation will follow in a later section.

The lesson begins with the teacher reading the “Mega Mining Mart” introduction. The introduction is read using the skip reading technique. Skip reading is when the teacher reads aloud from a passage and stops at important words or phrases. Students, as a class, then say the word or phrase, and the teacher resumes reading. After the introduction is finished, the teacher discusses data mining (with relevant examples) and its relation to the upcoming activity.

After the discussion on data mining, the teacher directs the students’ attention to the software engineering career sheet. Designed and written by the Fellows, this sheet provides all of the necessary information and salient facts about the field of software engineering. In the event that students have questions, additional material is provided that the teacher may reference.

Students are then directed to the “Mega Mining Mart” packet where they will begin their store design by spending up to 10 minutes brainstorming categories for the food and sundry items provided to them.

After their categories have been determined, students complete the store layout activity by organizing the items in their store. It is important that students keep in mind their store categories as well as the data mining associations when working on their layout. The teacher circulates throughout the room and checks student work. If it meets the requirements, students are allowed to permanently place their items (using clear tape) on their store layout.

After students have completed their store design, they move on to the chemistry questions located in the packet. These questions asked the students to identify trends in their store. Several questions asked students where they would place a new item (one not on their original list) in their store and to justify their choice. The Fellows made the connection between a new item being added and a new element being discovered. Students were shown that the elements in the periodic table are arranged systematically and according to trends such as increasing atomic weight and increasing number valence electrons.

Students were also shown a version of Mendeleev’s original periodic table that included blank spaces for 10 elements that Mendeleev thought existed but did not know what they were. Students were asked to relate these blank spaces back to their store and postulate how an empty space would be represented in their layout.

When all of the students have finished their packets, the teacher leads a discussion that compares their store layout to the periodic table. The main point of this discussion is to make sure the students understand the ideas of trends and correlations as they relate back to the periodic table.

**ASSESSMENT**

A checklist that describes student learning assessment was developed by the Fellows. The assessment was based on class participation, class discussion, store layout design, and questions asked in the “Mega Mining Mart” packet. In addition, the Fellows designed a grading rubric that addresses the four main components of the lesson – participation, packet questions, store layout, and professional presentation (appearance) of work [7]. This grading rubric divided each of the four main components into five possible grades – excellent, good, mediocre, poor, and incomplete. The four main components and five possible grades formed a grid. Inside the cells of the grid, the Fellows indicated exactly what was required of the student in order to receive that grade in the specified area. The grading rubric was distributed to the students at the beginning of the lesson, and the Fellows made sure to draw the students’ attention to the grading rubric so they were clear as to exactly what was expected from them.

Assessment of the lesson is achieved via a student feedback form. Students completed the form at the end of the lesson and submitted it to the Fellows. The Fellows also actively watched the students during the lesson and made notes on their interest level.

**OUTCOMES**

Outcomes for “Mega Mining Mart” are divided into two categories – lesson grades and feedback from students. The lesson was implemented by the Fellows in two physical science classes with the same secondary educator. This section describes the outcomes of student grades and feedback.

I. Grades Summary

Student grades for the lesson fell above average grades for the class. The average score (out of 50 points possible) for the classes was 38.92 (77.8%) and 41.89 (83.9%). The average score for a typical lesson in these classes was 59.7% and 68.2% respectively. The lowest score was a 28 (56.0%), and the majority of scores fell between 35 (70.0%) and 47 (94.0%) with a mode grade of 43 (86.0%). Unfortunately, a more rigorous statistical comparison is not possible because we do not have access to additional details about student grades.

In order to ensure accurate and fair grading, the Fellows used a specialized grading rubric (described in the section above) that allowed them to assign specific point values to the different aspects of the lesson.

II. Student Feedback

The anonymous feedback provided by students is important for the assessment of the lesson and, more generally, for judging how well the lesson meets the goals of Project STEP. Students complete one-page sheet which asks several questions about the lesson and engineering, and they are also able to leave free-form comments that describe their likes.
and dislikes. Table 1 summarizes several questions and the results.

<table>
<thead>
<tr>
<th>TABLE I STUDENT FEEDBACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>How has this lesson affected your interest in engineering?</td>
</tr>
<tr>
<td>Increased</td>
</tr>
<tr>
<td>No Change</td>
</tr>
<tr>
<td>Decreased</td>
</tr>
<tr>
<td>How much did you learn in this lesson?</td>
</tr>
<tr>
<td>A Lot</td>
</tr>
<tr>
<td>Some</td>
</tr>
<tr>
<td>Nothing New</td>
</tr>
<tr>
<td>Did this lesson make you more confident about learning science?</td>
</tr>
<tr>
<td>Definitely</td>
</tr>
<tr>
<td>Some</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Not Sure</td>
</tr>
</tbody>
</table>

The results shown in Table 1 are encouraging. The majority of students displayed an increased interest in engineering. All students reported that they learned something new from the lesson, and a majority of students (81.48%) of students felt more confident about their ability to learn science.

Students gave their feedback in two categories – what they liked the most and the least about the lesson. Student responses varied, but the majority of student responses were positive and encouraging. Several students commented that they were not aware of engineering being used so much in things they see and use everyday and how this was “cool.” Students also indicated that they enjoyed the discussion about software engineering and that they were now considering that a potential path of study when entering college.

In addition to the engineering and science aspects, students also commented on the ability to exercise their creativity. They commented that they enjoyed the ability to make decisions and arrange their store layout so that the trends were apparent and made sense to the consumer, and the use of manipulatives to arrange their store layout made it easier to work collectively in a small group.

Students reported that they never thought of a grocery store in terms of trends and relationships, and that they believed they had a better understanding of the arrangement of elements on the periodic table.

OBSERVATIONS AND REFLECTIONS

Throughout the design, development, and execution of the “Mega Mining Mart” lesson, the Fellows recorded their observations. After students completed the lesson, the Fellows recorded their reflections about the effectiveness of the lesson and possible changes for future implementation.

In one of the classes, students were given an unlimited amount of time to decide on the categories for their grocery store items. This proved to be an inefficient use of classroom time. Since they were working in groups of two or three, the authors observed that a large amount of time was wasted by the students due to endless disagreement about the categories and what item fit into which one. A solution to this was to impose a strict 10 minute time limit on the category creation process. When this was enacted for the second class, the students were able to focus more on the task and complete it more efficiently and effectively. However, even with this time limit, the students from the second class finished the same amount of the lesson in day one than the first class. The students in the second class required more individual guidance than the first class, and this caused them to complete the same amount of work in a larger amount of time.

The two most significant impacts of the lesson were a greater understanding of trends in the periodic table and a heightened interest in engineering as it applies to everyday items and situations. Students commented to the Fellows that they understood why the periodic table was arranged the way it is, and they felt confident they could explain this trend to others and successfully answer questions about trends on an exam. Students also became more interested in engineering and data mining in their everyday lives. On subsequent visits to the classroom, the Fellows were asked several questions about data mining and how it can potentially impact their privacy. This was not a topic of discussion during the lesson, and it shows that students were taking the knowledge they learned, thinking about it, and making a logical extension. Students also asked pointed questions about the technologies used for data mining, and they were curious to know more about software engineers.

CONCLUSIONS

This paper illustrates how advanced computing topics (such as data mining) can be successfully integrated into high school curriculum without interfering with the necessity to meet state and national academic content standards. Through this lesson and many like it, Project STEP is making inroads into secondary education by providing hands-on, inquiry-based lessons that are based on engineering principles and relevant to students’ lives. The implementation of activities and the presence of Fellows in the classroom have resulted in increased engagement of high school students with STEM subjects and increased confidence in their abilities to learn these subjects. Teachers have also benefited from exposure to engineering topics that broaden the applications of their disciplines. Details about this lesson, other lessons, and the achievements of Project STEP can be found online [1]. In addition to the STEP website, the lesson has been disseminated on the Ohio Resource Center’s website [6] for educators. The ORC has rigorous reviewing standards to ensure the quality and completeness of lessons posted there.

The Mega Mining Mart lesson is currently being adapted for use in a related effort that supports lesson delivery by undergraduate students in CPS high school STEM classrooms [5]. Since the focus of this NSF-funded effort is on computing-related topics in order to broaden participation in computing, the chemistry aspect has been replaced with extended material on data mining and pattern recognition. Prior to the Mega Mining Mart hands-on

Session T1A

October 22 – 25, 2008, Saratoga Springs, NY

38th ASEE/IEEE Frontiers in Education Conference

T1A-5
activity with grocery store layout design, the students are given a small data set representing customer purchases. They learn to write association rules by hand, then discuss an optimized set of rules created by a data mining algorithm. The students verify by hand that the rules hold for the given data set. Following the creation and application of association rules as a logical exercise, the students do the Mega Mining Mart grocery store activity as a practical application of data mining’s influence and impact on our daily lives.

ACKNOWLEDGMENT

This work is supported by National Science Foundation grants #0139312 and #0634473 and matching funds from the University of Cincinnati. The authors gratefully acknowledge Debra Hutchinson for allowing us to work with her students.

REFERENCES


AUTHOR INFORMATION

Bartley Richardson, Graduate Student (PhD) and Previous Fellow, University of Cincinnati, Department of Computer Science, richarbe@ececs.uc.edu

Karen Davis, PhD, Associate Professor, University of Cincinnati, Department of Electrical and Computer Engineering, karen.davis@uc.edu

Michelle Daniel Beach, Secondary Science Teacher and Previous Fellow, Midpark High School, Middleburg Heights, OH, mbeach@berea.k12.oh.us