Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #46109

Data Mining Application in an Introductory Engineering Physics Lab

Prof. Rodrigo Cutri, Maua Institute of Techonology

Cutri holds a degree in Electrical Engineering from Maua Institute of Technology (2001), MSc (2004) and Ph.D. (2007) in Electrical Engineering - University of SÃ£o Paulo. He is currently Titular Professor of Maua Institute of Technology, Professor of the

Dr. Octavio Mattasoglio Neto, Instituto Mauá de Tecnologia

Undergraduate in Physics (1983), Master in Science (1989) and Phd in Education (1998) all of them from Universidade de São Paulo. Professor of Physics at Mauá Institute of Technology, since 1994 and President of Teacher's Academy at the same Institution.

Dr. Nair Stem, IMT

- Graduated at Physics (Bachelor) at IFUSP, Master at Electrical Engineering and Doctor at Electrical Engineering at EPUSP.

Data Mining Application in an Introductory Engineering Physics Lab

Abstract

This study explores the application of data mining techniques in Physics laboratories for Engineering, aiming to enhance the educational process and students' understanding of physical phenomena. The primary objective is to analyze how the use of Orange Data Mining software can facilitate the analysis of large volumes of experimental data, enabling students to identify patterns and extract relevant insights for their investigations. The adopted methodology involved conducting Physics experiments in non-conservative systems, where students collected data on friction across different materials and utilized techniques such as linear regression and clustering (K-means) to analyze the results.

Following the application of these techniques, a pre- and post-class knowledge assessment was conducted using a Likert-type questionnaire. The results indicated a significant improvement in participants' understanding of concepts related to force, energy, and data analysis. Additionally, the experience with Orange provided students with greater familiarity with data science tools, better preparing them for the challenges of the engineering job market.

This study highlights the importance of integrating data mining techniques into engineering education, offering an innovative approach to learning complex physical concepts.

Introduction

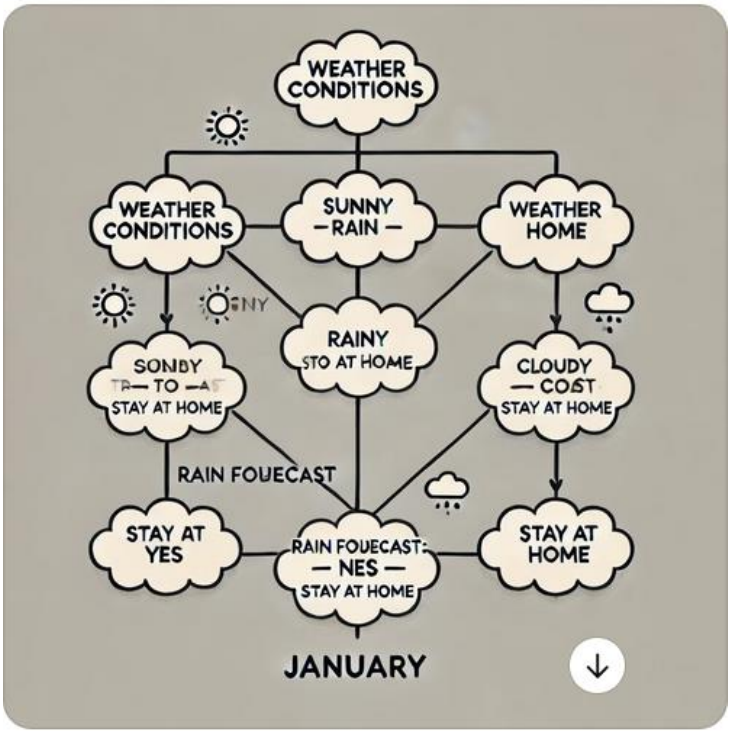
This complete paper explores the application of data mining techniques in Physics laboratories for Engineering, aiming to enhance the educational process and students' understanding of physical phenomena. The primary objective is to analyze how the use of Orange Data Mining software can facilitate the analysis of large volumes of experimental data, enabling students to identify patterns and extract relevant insights for their investigations.

The advancement of computational technologies has profoundly transformed the landscape of Engineering education, especially regarding the application of data mining in Physics laboratories. Data mining is a set of techniques that allows for extracting useful knowledge from large volumes of data, enabling insights that would otherwise be inaccessible. In the context of Physics laboratories for Engineering, these techniques are fundamental for managing the growing complexity of experiments and the volume of collected data.

In the Physics lab, various data mining techniques are applied, each serving a specific purpose:

Decision Trees: Used to create predictive models from a dataset, decision trees are particularly useful for classifying conditions and predicting outcomes based on input variables. For example, in a weather conditions experiment, a decision tree could be used to predict whether certain weather conditions indicate the feasibility of outdoor activities or not.

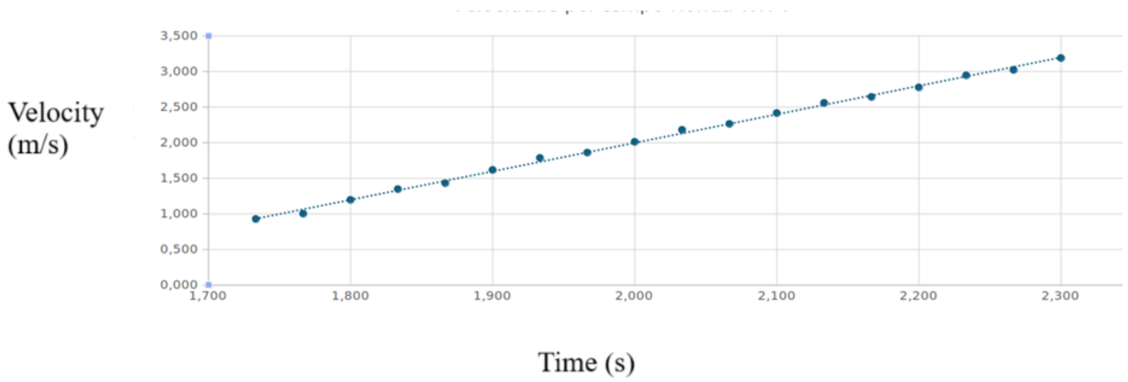
Figure 1: Example of weather conditions´ decision tree to decide whether stay home or not



Source: Chat Gpt

Linear Regression: This technique is essential for predicting continuous values from independent variables. In Physics experiments, such as studying the relationship between speed and time, linear regression can be used to fit a function to experimental data and predict future behaviors, such as the speed of an object at a specific moment.

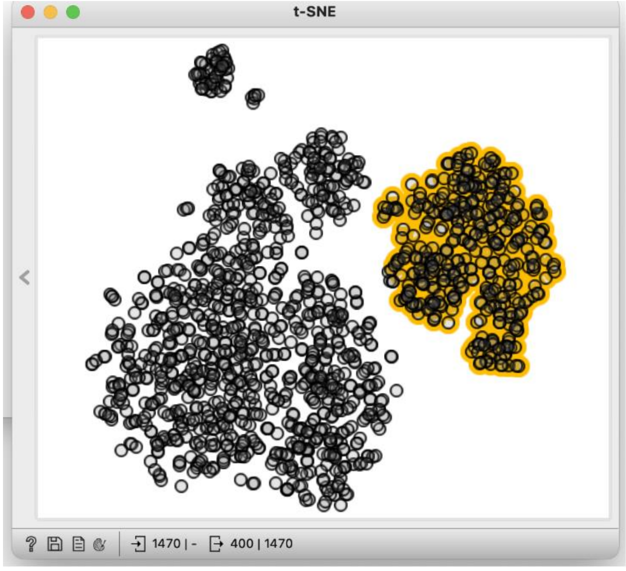
Figure 2: Velocity of Honda WR-V car as function of time



Source: the authors

Clustering: Clustering groups similar data together, facilitating the identification of hidden patterns in experiments. In an engineering context, clustering can be used to analyze friction between different materials on a surface, grouping similar data to determine which materials offer the least resistance to movement.

Figure 3: Examples of clusterings using Orange Data Mining



Source: https://orangedatamining.com/blog/characterizing-clusters-with-a-box-plot/

Orange Data Mining

The software Orange stands out as an intuitive and powerful tool for performing data mining analyses in physics laboratories. With a user-friendly graphical interface, Orange allows students and researchers to conduct complex analyses without the need for advanced programming skills, making it ideal for engineering education. In addition to implementing algorithms like decision trees, linear regression, and clustering, Orange offers interactive visualizations that make it easier to understand the results.

The Orange software provides an intuitive platform for loading experimental data, applying clustering algorithms like K-means, and uncovering patterns that highlight specific characteristics to be analyzed. This approach not only makes data more comprehensible but also equips students with the ability to make informed decisions based on data analysis.

Data Analysis Proposal: Using Clusters for Physical Data Interpretation with Orange

The adopted methodology involved conducting Physics experiments in non-conservative systems, where students collected data on friction across different materials and used techniques such as linear regression and clustering (K-means) to analyze the results.

Objectives of the Proposal

1. Understand the Fundamentals of Clustering: Introduce students to clustering concepts, emphasizing their applications in identifying patterns and grouping similar data points.

2. Apply K-means Algorithm: Use Orange to cluster experimental physical data, visualizing relationships and trends within datasets.

3. Enhance Decision-Making Skills: Enable students to interpret clustered data for practical insights, promoting data-driven decision-making.

Proposed Activities

1. Introduction to Clustering and K-means:

o Theoretical explanation of clustering algorithms, focusing on K-means.

o Discussion on the importance of clustering in physical sciences and data analysis.

2. Loading Experimental Data into Orange:

o Guide students to upload and preprocess physical datasets in Orange.

o Explain the significance of data normalization and feature selection in clustering.

3. Applying the K-means Algorithm:

o Demonstrate how to set the number of clusters and run the K-means algorithm in Orange.

o Visualize results using scatter plots, dendrograms, and other tools within Orange.

4. Interpreting Clustering Results:

o Analyze identified clusters to determine patterns and characteristics (e.g., similar measurements or conditions within experimental setups).

o Discuss practical implications of the clusters in a real-world context.

5. Student Projects:

o Assign students to design their own analysis using experimental physical data.

o Evaluate how well students apply clustering techniques and interpret results.

Benefits of the Approach

- Hands-on Experience: Students gain practical skills in clustering and data visualization.

- Interdisciplinary Learning: Links physical sciences with data science, broadening student expertise.

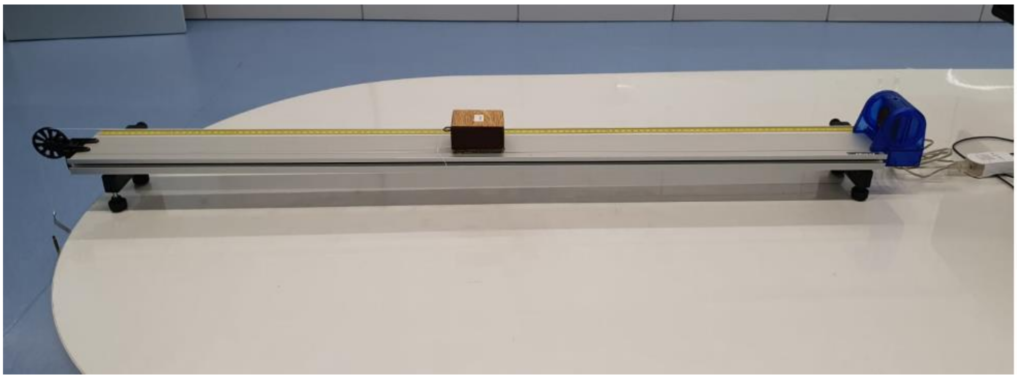
- Enhanced Comprehension: Patterns and relationships in data become more accessible, fostering deeper understanding.

- Problem-Solving Skills: Students learn to approach physical data challenges using computational tools and statistical techniques.

The experiment

In this experiment, students have at their disposal a block with a carpet of mass mB on a horizontal steel track. Another option is to use a block with EVA. An ideal string (negligible mass and inextensible) connects this block to a suspended body with a known mass Ms through an ideal pulley. The pulley has negligible mass, and we will assume it is friction-free. Students must model the problem to determine the coefficient of kinetic friction when the system is released.

The modeling should be done using two different methods: one based on Newton's laws and the other on energy conservation.



From the data generated by each group, a large database is created with measurements of EVA and carpet sliding on the steel track (Pasco track). The data were scrambled without identification in Excel spreadsheet. Using the K-means algorithm, students must group similar data and analyze the work done by friction for both groups. The friction coefficients were calculated by both methods. Thus, they should:

a) Identify the cluster that likely corresponds to the carpet and EVA;

b) Determine the relationship between Fat and Wfat;

c) Determine which cluster shows greater energy loss;

d) Decide which material the student would choose to cover their steel track.

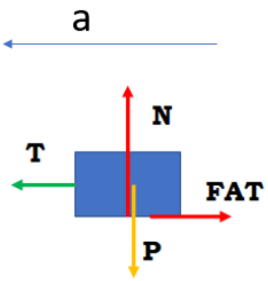
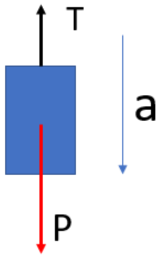
Physical Modeling

Students engaged in the study of two modeling approaches:

1º) Using Newton's Second Law

This approach involves applying Newton's Second Law to analyze and model physical systems, focusing on the relationships between forces, mass, and acceleration.

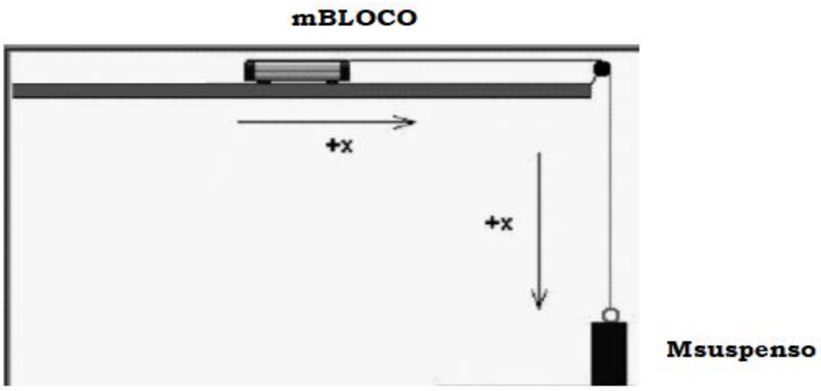
$$\sum \vec{F} = m\vec{a}$$

 Direction x: T-Fat=m_B a Direction y: N-P=0	 T-P=m_S(-a)

2º) Using the Principle of Mechanical Energy Conservation

This approach applies the principle of mechanical energy conservation to model physical systems, emphasizing the relationship between kinetic and potential energy and their transformations within a system.

$$W_{fat} = E_{mB} - E_{mA}$$



Initially, the system is at rest, with the body on the table at a height h' relative to the indicated reference point. The suspended body hangs at a height x.

After body in plane moves a distance x, and the suspended body also descends by x, both bodies will have gained velocity, thus acquiring kinetic energy.

The suspended body will reach the reference point at zero height, while the body on the table will remain at height h' .

This setup demonstrates the conversion of potential energy into kinetic energy for both bodies, with the suspended body's height reducing to the reference point while the other body's elevation remains unchanged. This dynamic interaction exemplifies the principles of both energy conservation and motion within the system.

$$W_{fat} = |Fat||d|\cos 180^{o}$$

$$W_{fat} = E_{mB} - E_{mA}$$

$$E_{mA} = M_s g x + m_B g h'$$

$$E_{mB} = mBgh' + (1/2)\ \text{mB v\^{}2} + (1/2)\ M_s\ \text{v\^{}2}$$

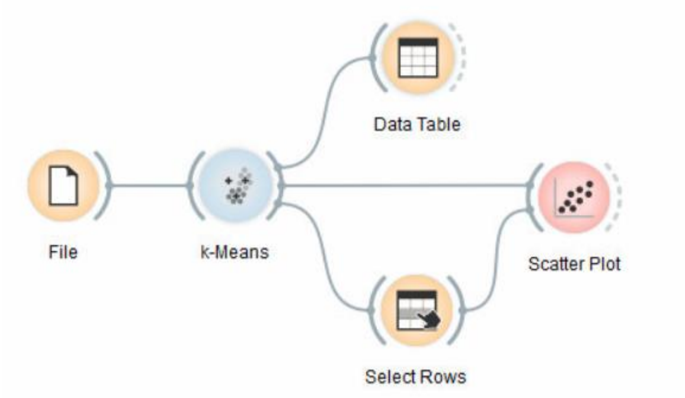
After collecting data, students are tasked with comparing the two modeling methods (Newton's Second Law and the Principle of Energy Conservation) by calculating the respective percentage error for each method.

Dataset Development

Each group generates data during experiments involving EVA and carpet sliding on a steel track. These datasets are compiled into a comprehensive database. Using the K-means algorithm in Orange Data Mining, students are required to group similar data and analyze the work done by friction across both groups. The friction coefficients calculated using both methods are also included in the analysis.

Clusters – Orange Data Mining

The following structure was set up in Orange Data Mining, and it is up to the students to determine:



a) Identify the cluster that most likely corresponds to the carpet and EVA.

b) Determine the relationship between Fat (friction force) and Wfat (work done by friction).

c) Identify which cluster represents the greatest energy loss.

d) Determine which material the student would choose to cover their steel track.

Evaluation

A pre-and post-lesson knowledge assessment was conducted using a Likert-type questionnaire (appendix A). The experiment was conducted with approximately 340 students divided into groups of 3 to 4 in laboratory classes (100 minutes). Around 119 students responded to the questionnaire.

Subject	Question	Pre-class % Accuracy	Post-class % Accuracy	% Difference
Kinetic Energy and Friction	1	97%	97%	0%
Kinetic Energy and Friction	2	99%	97%	-2%
Forces and Friction	3	98%	96%	-2%
Forces and Friction	4	98%	97%	-1%

Forces and Friction	5	64%	64%	0%
Energy Conservation	6	92%	92%	0%
Energy Conservation	7	92%	92%	0%
Data Analysis	8	87%	91%	4%
Data Analysis	9	86%	93%	7%
Data Analysis	10	93%	93%	0%

key insights from the data:

1. Consistent Performance in Certain Topics

For Kinetic Energy and Friction (Question 1) and Energy Conservation (Questions 6 and 7), pre- and post-class accuracy remained unchanged at 97% and 92%, respectively. This indicates a solid initial understanding of these concepts among students and limited room for improvement during the class.

Insight: These topics may already be well-understood by students, or the activities did not significantly enhance their understanding.

2. Slight Decline in Accuracy

Kinetic Energy and Friction (Question 2) and Forces and Friction (Questions 3 and 4) saw a small drop in accuracy by 1-2% post-class.

Insight: This may suggest that the activities introduced additional complexity, potentially confusing some students. Further review or simplification of these concepts may be beneficial.

3. No Improvement in Weak Areas

For Forces and Friction (Question 5), pre- and post-class accuracy remained at 64%, indicating no improvement.

Insight: This topic likely requires more focused attention, as it appears to be a challenging area for students.

4. Significant Improvement in Data Analysis

Data Analysis (Questions 8 and 9) showed noticeable improvement, with increases of 4% and 7%, respectively, from pre- to post-class.

Insight: The activities and tools used in teaching Data Analysis, such as Orange Data Mining, appear to have effectively enhanced student understanding of these concepts.

5. Consistently High Accuracy

For Data Analysis (Question 10), accuracy remained at a high level of 93%, indicating strong comprehension both before and after the class.

Insight: Students are already confident in this topic, and the activities reinforced their existing knowledge.

General Observations

Strong Topics: Data Analysis and Energy Conservation concepts are generally well-understood and benefited from the class activities.

Weak Topics: Forces and Friction (especially Question 5) need more targeted interventions or alternative teaching methods.

Room for Improvement: While some topics saw improvement, the small declines in a few questions highlight the need to ensure clarity and reinforcement during class activities.

When analyzing the data from the previous table, it is noticeable that the percentages for pre- and post-class accuracy are very close, raising the hypothesis that students may have researched to answer the questionnaire. Therefore, the following questions were subsequently asked to the students:

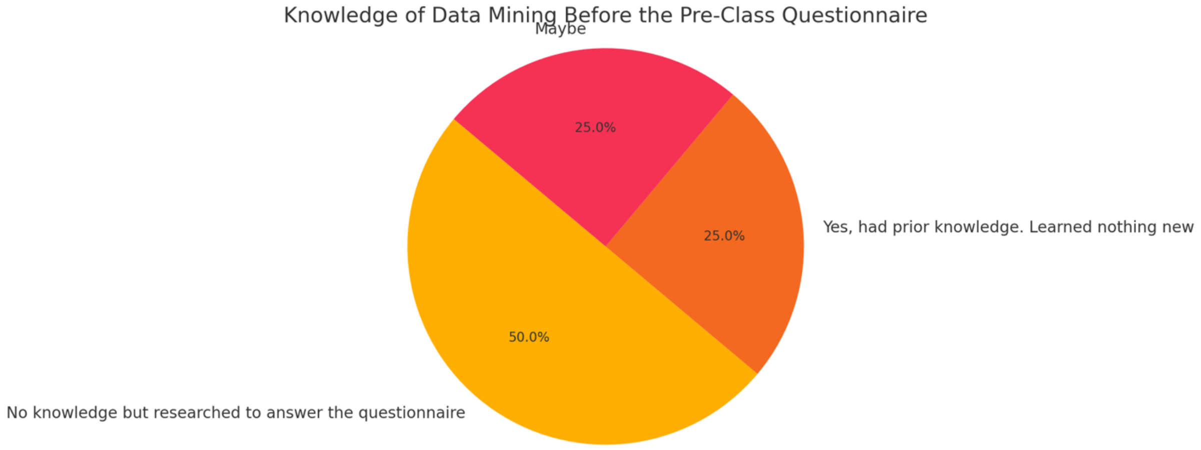
a) What was their knowledge before the pre-class?

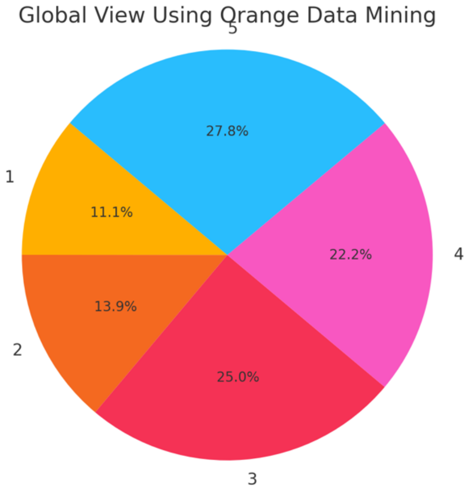
b) What is their overall opinion on the use of Orange?

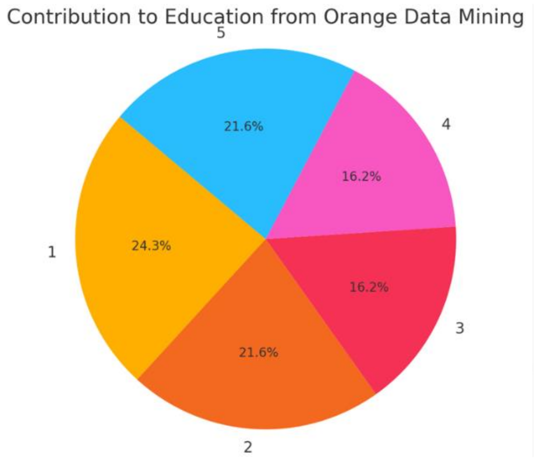
c) What is their perception of Orange's contribution to the learning process?

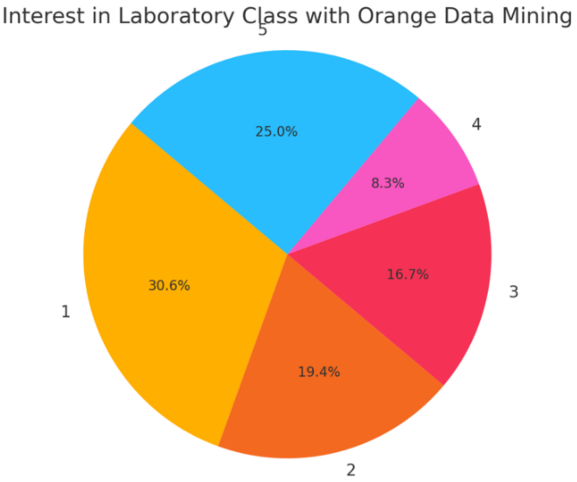
d) How did Orange increase their interest in the Physics laboratory?

The following graphs and analyses were obtained:


Knowledge of Data Mining Before the Pre-Class Questionnaire


Global View Using Orange Data Mining


Contribution to Education from Orange Data Mining


Interest in Laboratory Class with Orange Data Mining

Key insights obtained from the graphs:

Knowledge of Data Mining Before the Pre-Class Questionnaire

50% of respondents had no prior knowledge but researched online or through other means to answer the questionnaire.

Only 25% had prior knowledge and did not learn anything new.

Another 25% were undecided ("Maybe") about their knowledge.

Insight: The majority of students (75%) did not have a solid understanding of data mining before the class, indicating an opportunity to introduce basic concepts and engage them more deeply in the topic.

Interest in the Laboratory Class with Orange Data Mining

30.6% gave a score of 1, indicating low interest.

25% gave a score of 5, suggesting that a quarter of the students found the introduction of Orange very engaging.

Intermediate scores (2 to 4) accounted for 44.4%, showing moderate interest.

Insight: The introduction of Orange Data Mining divided opinions; while some students found the tool motivating, many did not show significant enthusiasm. Pedagogical adjustments may be necessary to engage more students.

Global View Using Orange Data Mining

27.8% of students gave a score of 5, indicating that the methodology significantly helped provide a global view of the concepts.

Scores of 4 and 5 combined accounted for 50%, suggesting that half of the students recognized the positive impact of the approach.

11.1% gave a score of 1, suggesting that a small portion did not perceive the benefit.

Insight: The methodology effectively provides a broader understanding of the concepts taught, but there is room to improve the experience for a minority who did not perceive the same benefits.

Contribution to Education

The responses were balanced, with 24.3% giving a score of 1 (lowest contribution) and 21.6% giving a score of 5 (highest contribution).

Intermediate scores (2 to 4) accounted for 54%, indicating a moderate perception of contribution.

> Insight: While a significant portion of students sees value in the tool, the variation in responses suggests the need for strategies to more clearly demonstrate the practical applicability of Orange Data Mining.

Final Considerations

Variable Engagement: While some students recognize the value of the tool, others still need to be convinced of its relevance.

Opportunity for Improvement: Adapting pedagogical strategies to make the use of Orange more practical and directly aligned with educational objectives can increase its impact.

Overall Positive Impact: Half of the students consider the tool to have contributed to a broader understanding of the concepts and to their academic development, reinforcing the importance of keeping it in the curriculum.

Additionally, in qualitative research with the students, the simple use of a data mining tool was important in sparking curiosity about the subject and helping them understand the importance of data processing. The results indicated a significant improvement in the participants' understanding of the concepts of force, energy, and data analysis. Moreover, the experience with Orange allowed students to become more familiar with data science tools, better preparing them for the challenges of the engineering job market. This study underscores the importance of integrating data mining techniques into physics analysis, offering an innovative approach to learning complex physical concepts.

The use of data mining and tools like Orange in teaching Physics for Engineering addresses a growing demand in the job market for professionals who not only master traditional engineering concepts but also possess skills in computational thinking and data analysis. Modern industries increasingly value engineers capable of handling large volumes of data, extracting insights, and applying this knowledge to optimize processes, innovate products, and solve complex problems.

In this context, the incorporation of data mining into the Engineering curriculum is not merely an adaptation to new technologies but an essential preparation for the future. Engineers of tomorrow will need to be familiar with tools like Orange, understand the fundamentals of data mining algorithms, and know how to apply these concepts in practical settings. These skills are not just desirable but indispensable in today's and tomorrow's job market.

In summary, teaching Physics for Engineering, enriched with data mining techniques and tools like Orange, provides students with a robust skill set that positions them ahead in a competitive and ever-evolving market.

Acknowledgments

We would like to express our sincere gratitude to the XXXXXX and its dedicated technical staff, with a special thanks to technician XXXXX.

References

[1] Agung Triayudi et al 2022 Data Mining K-Means Algorithm for Performance Analysis J. Phys.: Conf. Ser. 2394 012031 DOI 10.1088/1742-6596/2394/1/012031

[2] Agung Triayudi and Wahyu Oktri Widyarto Educational Data Mining Analysis Using Classification Techniques 2021 J. Phys.: Conf. Ser. 1933 012061 DOI 10.1088/1742-6596/1933/1/012061

[3] April Lia Hananto et al Analysis Of Drug Data Mining With Clustering Technique Using K-Means Algorithm 2021 J. Phys.: Conf. Ser. 1908 012024 DOI 10.1088/1742-6596/1908/1/012024

[4] Hussain, S., Atallah, R., Kamsin, A., Hazarika, J. (2019). Classification, Clustering and Association Rule Mining in Educational Datasets Using Data Mining Tools: A Case Study. In: Silhavy, R. (eds) Cybernetics and Algorithms in Intelligent Systems . CSOC2018 2018. Advances in Intelligent Systems and Computing, vol 765. Springer, Cham. https://doi.org/10.1007/978-3-319-91192-2_21

[5] B.M. Monjurul Alom, Matthew Courtney, "Educational Data Mining: A Case Study Perspectives from Primary to University Education in Australia", International Journal of Information Technology and Computer Science(IJITCS), Vol.10, No.2, pp.1-9, 2018. DOI:10.5815/ijitcs.2018.02.01

[6] Cutri, R., & Stem, N., & Mattasoglio, O. (2024, June), *The Physics of Gym Elastic: Elastic Force and Energy of a Non-Linear Material* Paper presented at 2024 ASEE Annual Conference & Exposition, Portland, Oregon. 10.18260/1-2--48126

Appendix A

Questions:

Kinetic Energy and Friction:

1. Imagine a wooden block sliding on a horizontal surface. Initially, the block moves at a constant velocity. What happens to the block's kinetic energy over time if the surface is perfectly smooth (frictionless)?
   a) The block's kinetic energy remains constant.
   b) The block's kinetic energy increases indefinitely.
   c) The block's kinetic energy decreases until the block stops.
   d) The block's kinetic energy oscillates between maximum and minimum values.
   Correct answer: a) The block's kinetic energy remains constant.

2. Imagine a wooden block sliding down an **INCLINED PLANE**. Initially, the block starts from rest at the highest point of the plane. What happens to the block's kinetic energy over time if the surface is perfectly smooth (frictionless)?
   a) The block's kinetic energy remains constant.
   b) The block's kinetic energy increases up to the end of the track.
   c) The block's kinetic energy decreases until the block stops.
   d) The block's kinetic energy oscillates between maximum and minimum values.
   Correct answer: b) The block's kinetic energy increases up to the end of track.

Forces and Friction:

3. A 20 kg box is at rest on a horizontal table. What is the approximate value of the normal force exerted by the table on the box?

a) 0 N

b) 10 N

c) 20 N

d) 200 N

Correct answer: d) 200 N

4. A wooden block with mass M slides on a horizontal surface with an initial velocity v0. The coefficient of kinetic friction between the block and the surface is μ. What will happen to the block, and why?

   a) The block will continue moving at a constant velocity because no forces are acting on it.

   b) The block will gradually decelerate and stop due to the kinetic friction force acting against its motion.

   c) The block will gradually accelerate until reaching a maximum velocity due to the kinetic friction force acting in favor of its motion.

   d) The block will oscillate between points of rest and motion due to the kinetic friction varying over time.

   Correct answer: b) The block will gradually decelerate and stop due to the kinetic friction force acting against its motion.

5. An experiment is conducted to determine the relationship between the force applied to a spring and its displacement. The data collected during the experiment is presented in the following table:

Force (N)	**Displacement (m)**
10	0.1
20	0.2
30	0.3
40	0.4
50	0.5

What is the spring constant?

a) 500

b) 100

c) 50

d) I don’t know

Correct answer: b) 100 N/m

Conservation of Energy:

6. Imagine a roller coaster in an amusement park. At a certain point on the track without friction, the cars reach their maximum speed. At this moment, what type of energy do they have?

a) Only gravitational potential energy.

b) Only kinetic energy.

c) A combination of gravitational potential energy and kinetic energy.

d) The mechanical energy of the cars dissipates due to friction with the track and is not conserved.

Correct answer: b) Only kinetic energy.

7. Imagine a toy car being released from the top of a frictionless slope. During the descent, what type of energy does the car have?

   a) Only gravitational potential energy.

   b) Only kinetic energy.

   c) A combination of gravitational potential energy and kinetic energy.

   d) The mechanical energy of the car dissipates in the air and is not conserved.

   Correct answer: c) A combination of gravitational potential energy and kinetic energy.

Data Analysis:

8. An experiment is conducted to determine the relationship between the force applied to a spring and its displacement. The data collected during the experiment is presented in the following table:

Force (N)	**Displacement (m)**
10	0.1
20	0.2
30	0.3
40	0.4
50	0.5

Which data science tool would you use to analyze this data and determine the relationship between the force and the spring's displacement?

a) Linear regression

b) K-means (Clustering)

c) Decision trees

d) I don’t know

Correct answer: a) Linear regression

9. Imagine a researcher collecting data on customer purchasing behavior in a supermarket. Each customer is represented by a set of characteristics such as age, gender, income, and purchased items. The researcher's goal is to group customers into segments based on their purchasing characteristics, identifying behavioral patterns. What data analysis method should they use for this task?

   a) Linear regression

   b) K-means (Clustering)

   c) Decision trees

   d) I don’t know

   Correct answer: b) K-means (Clustering)

10. Imagine a doctor diagnosing a patient based on their symptoms. The doctor uses a set of decision rules to assist in the diagnostic process. Each rule involves evaluating a symptom and leads to one of the following possibilities: ordering more tests, referring the patient to a specialist, or concluding the diagnosis. What data analysis method can be used to represent these decision rules intuitively and assist the doctor in diagnosis?

    a) Linear regression

    b) K-means (Clustering)

    c) Decision trees

    d) I don’t know

    Correct answer: c) Decision trees