Data Analytics for Decision Making at Academic Departments

Dr. Ashwin Satyanarayana, New York City College of Technology

Dr. Ashwin Satyanarayana is currently the Chair and Associate Professor with the Department of Computer Systems Technology, New York City College of Technology (CUNY). Prior to this, Dr. Satyanarayana was a Research Scientist at Microsoft in Seattle from 2006 to 2012, where he worked on several Big Data problems including Query Reformulation on Microsoft's search engine Bing. He holds a PhD in Computer Science (Data Mining) from SUNY, with particular emphasis on Data Mining and Big data analytics. He is an author or co-author of over 35 peer reviewed journal and conference publications. He has four patents in the area of Search Engine research. He is also a recipient of the Math Olympiad Award, and has served as Chair of the ASEE (American Society of Engineering Education) Mid-Atlantic Conference in 2018-2019. He also serves as an NSF (National Science Foundation) panelist.

Data Analytics for Decision Making at Academic Departments

Abstract

In the era of big data where data is being embraced by academic institutions, each academic department has access to lots of data –enrollment data, retention data, student outcomes, faculty productivity, student success rates and resource allocation. As a large four-year public institution, our institution serves a diverse student body where more than 60% of students are considered as economically disadvantaged. In our department (comprising 1728 students and 128 faculty), we are currently using data-driven decision-making to gain deeper insights into the needs of students, faculty and staff. Such planned and implemented data-driven strategy has transformed those insights into student success – retention and enrollment. Another area that data-driven culture has benefitted is in creating an unbiased environment (between faculty-student, administration-faculty, and chairperson-faculty), where collaboration and communication has become easier.

The main objective of this paper is to present our three data-analytic strategies: predictive, descriptive and prescriptive and how they have improved student outcomes, intervened at-risk students, strategized cost cutting in the department, projected actual outcomes and finally in determining the effectiveness of our data-decisions. For example, our predictive tool is helping identify potential low performing students at the course level and assigning them to mentoring and tutoring resources. Our prescriptive tool is helping with strategies and suggestions for cost-cutting and improving retention at the department level. Our descriptive tool is helping with data-driven unbiased communication between staff, faculty and students at the college level by giving data based feedback.

1. Introduction and Related Work

The last decade has seen an increase in the use of data analytics in various fields. Data analytics can be used in academic departments for decision making by analyzing large amounts of data to identify patterns, trends, and insights [1]. This information can be used to inform decisions related to resource allocation, course offerings, student performance, faculty productivity, and many other aspects of departmental operations. For example, data on student enrollment, academic performance, and demographics can be used to make informed decisions about what courses to offer, how to allocate resources to support student success, and how to attract and retain students. Additionally, data on faculty research productivity, grant funding, and teaching evaluations can be used to evaluate and reward the performance of individual faculty members and make decisions about departmental staffing and resource allocation.

Academic departments can use a variety of predictive, prescriptive, and descriptive analytics tools to inform decision making. We briefly introduce and survey the tools below.

1. Predictive Analytics Tools:
   These tools use statistical and machine learning algorithms to make predictions about future outcomes based on historical data. For example, a department might use predictive analytics to forecast future student enrollment or to identify students who are at risk of dropping out based on their previous academic performance. Nghe et. al. [2] compare the accuracy of decision tree and Bayesian network algorithms for predicting the academic performance of undergraduate and postgraduate students at two very different academic institutes: Can Tho University (CTU), a large national university in Viet Nam; and the Asian Institute of Technology (AIT), a small international postgraduate institute in Thailand that draws students from 86 different countries. Hamsa et al. [3] develop a student's academic performance prediction model, for the Bachelor and Master degree students in Computer Science and Electronics and Communication streams using two selected classification methods; Decision Tree and Fuzzy Genetic Algorithm.

2. Prescriptive Analytics Tools:
   These tools go beyond prediction and provide recommendations or decisions to achieve specific goals [4]. For example, a department might use prescriptive analytics to determine the optimal allocation of resources to improve student outcomes or to identify the most effective teaching methods based on student learning data.

3. Descriptive Analytics Tools:
   These tools summarize and describe data, often through visualizations and reports, to provide insights and understanding. For example, a department might use descriptive

analytics to gain insights into student demographics, enrollment trends, or faculty research productivity [5].

These tools can be implemented using a variety of software and platforms, including business intelligence (BI) tools, data visualization software, and data science platforms. The choice of tool depends on the specific needs of the department, the type of data being analyzed, and the goals of the analysis. However, some commonly used tools in academic departments include:

1. Data visualization tools [6]: These tools, such as Tableau and QlikView, allow users to explore and visualize data in an interactive way, making it easier to identify trends, patterns, and relationships in the data.
2. Predictive modeling tools [7]: These tools, such as R, SAS, and Python, allow users to build predictive models that can be used to make predictions about future outcomes based on historical data.
3. Machine learning algorithms [8]: These algorithms, such as decision trees, random forests, and support vector machines, can be used to build predictive models and to identify patterns and relationships in the data.
4. Descriptive analytics tools [9]: These tools, such as SAS and R, allow users to perform descriptive analysis, such as clustering and association analysis, to identify patterns and relationships in the data.
5. Data warehousing and ETL (extract, transform, load) tools [10]: These tools, such as Hadoop and Informatica, are used to store, manage, and process large amounts of data, making it easier for users to perform data analysis and to make informed decisions.

2. Motivation

Our institution is a senior urban college called New York City College of Technology (part of City University of New York (CUNY)), and ranks among the top 10 United States colleges that promote upward economic mobility [11]. With over 16,000 students as of Fall 2022, the college is home to one of the most diverse student bodies in the country. 58% of students reside in households with incomes of less than $40,000, and 70% of incoming first-year students receive financial aid. Many students also come to our institution with significant academic and intervention needs after attending underperforming schools that do not adequately prepare them for the academic rigor of college.

One of the main challenges (especially after the pandemic) being recruiting and enrolling new students, community and senior colleges (such as CUNY) also face a higher rate of attrition than traditional institutions. Many students attending these types of institutions face pressure from other factors such as full-time jobs, family obligations, transportation challenges and more. With the goal of retention, resource allocation and better enrollment strategies, in this paper, we use

analytics to provide strategies for struggling students to improve the quality of education at such two and four year colleges. One of the aims of this paper is to show how analytics allows the institution to make better predictions around students' potential for degree completion and retention. We also show how the academic chairpersons and administrators can have better, more engaging, unbiased, data-driven interactions with faculty when indicators flag negative performance regarding their teaching.

In this paper we will focus on a specific department within the institution - Computer Systems Technology (CST) department, which comprises 108 adjunct faculty and 20 full time faculty and had an enrollment of 1728 students in Fall 2022.

3. Research questions:

In this paper, we will try to address the following questions:
1. (Descriptive Analytics) What specific insights can we give faculty based on student evaluations, peer evaluations and grade distributions?
2. (Predictive Analytics) How effective is it to predict grades of students and how can we use that information to prescribe resources to students?
3. (Prescriptive Analytics) Can analytics prescribe cost-cutting and budgeting strategies at the institution level based on years of prior data?

4. Descriptive tools

4.1 Methodology

In this section, we demonstrate that by aggregating and summarizing data, descriptive analytics can provide a comprehensive view of the performance of individual faculty members and highlight any outliers or individuals who are not performing as well as others in their group [12]. We use a simple, yet popular data mining algorithm called *k-means* clustering to identify faculty who are flagged. Clustering is a data analysis technique that can be used to group data points based on their similarity [13]. In our work, we used clustering to group and evaluate teaching of faculty members based on three factors: (a) student evaluations, (b) peer faculty observations and (c) grade distributions (see Fig 1).
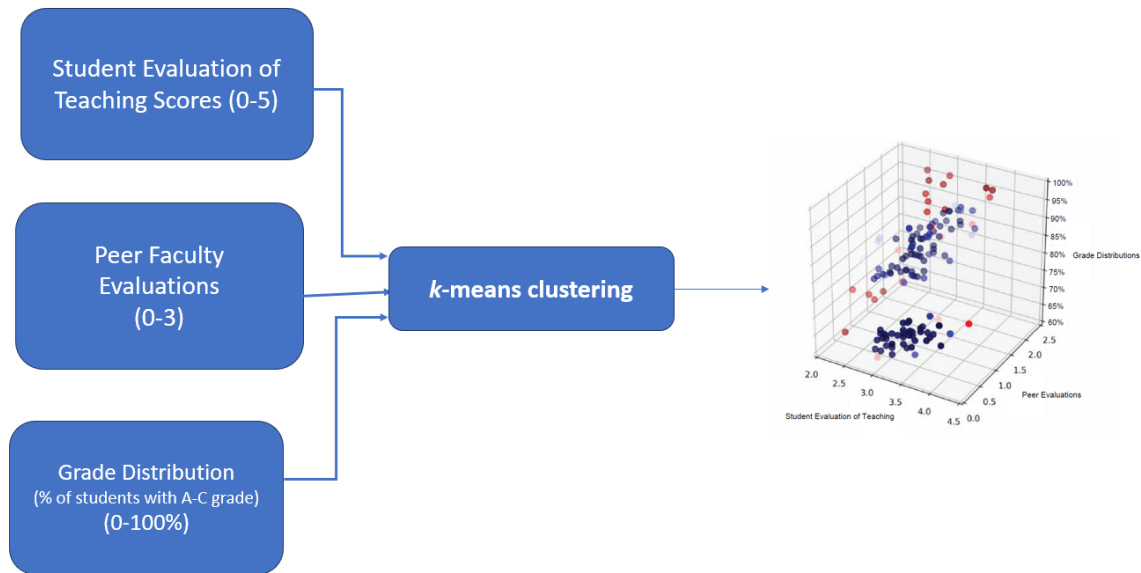
Fig 1: Three sources of data which are input to our descriptive tool for flagging faculty: (a) Student evaluation of teaching, (b) Peer evaluation of teaching and (c) Grade distribution by faculty.

This output provided by the descriptive tools is then used to flag those individuals for further review and provide them with support and resources (for e.g. summer teaching workshops) to help improve their performance. Additionally, we have seen that clustering helps identify any systematic issues or patterns in the data that may be contributing to poor performance, such as an instructor's communication skills of instructor, instructor availability to meet after classes, grading system, etc. [14].
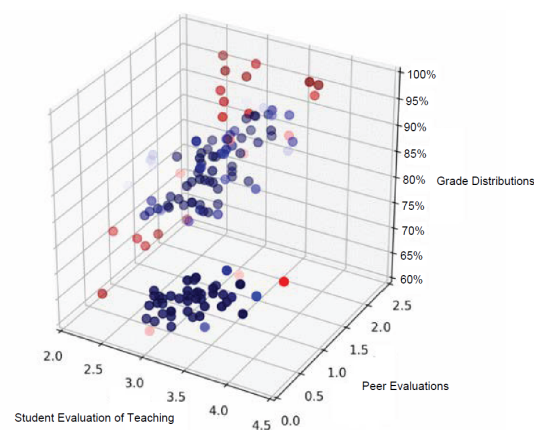

Fig 2: Output of the descriptive visualization tool, which flags 16 professors (in red) in Fall 2022.

4.2 Results:

We used the tool in Fall 2020 for the first time when it flagged 18 faculty (out of 120 faculty) with specific reasons for flagging them. The faculty appreciated and took the unbiased data driven feedback positively and 13 showed improvement by Fall 2021. In the latest usage of the tool in Fall 2022, the tool found 16 faculty as outliers (see Table 1).

Semester	Number of faculty identified as low performing	Percent of faculty who showed improvement after one year using the 3 performance metrics
Fall 2020	18	13 (72%)
Fall 2021	22	15 (68%)
Fall 2022	16	N/A

Table 1: Results showing improvement in quality of teaching of faculty who were flagged one year ago.

5. Predictive analytic tools:

Predictive data analytic tools can be used in predicting faculty and student performances by using statistical and machine learning algorithms to make predictions about future outcomes based on historical data.

5.1 Methodology

We used predictive data analytics to predict the grade of students based on data collected on previous academic performance, attendance, and demographic information to train a machine learning model to predict future grades for individual students. This information can then be used to identify students who are at risk of underperforming and provide them with targeted support and resources to help improve their grades. For each student taking a class, we use a survey questionnaire within the first month of classes consisting of general background questions and computer background questions. Some sample questions are as follows (more details are mentioned in our prior work in [15]): general questions such as - "Do you intend to enroll for an advanced degree when, or if, you complete your undergraduate degree?", "Do you have access to a computer where you live or work, or nearby that you can use for school work?", etc. For specific course related questions we have (for an introductory CS class): "Used computer to

analyze data (statistics, forecasting, etc.)?", "Developed Web pages or multimedia presentations?", etc.

The tool we are using is an ensemble classifier which comprises three classification algorithms (Random forest, Random trees and J48) to predict the grades of students from A to F.
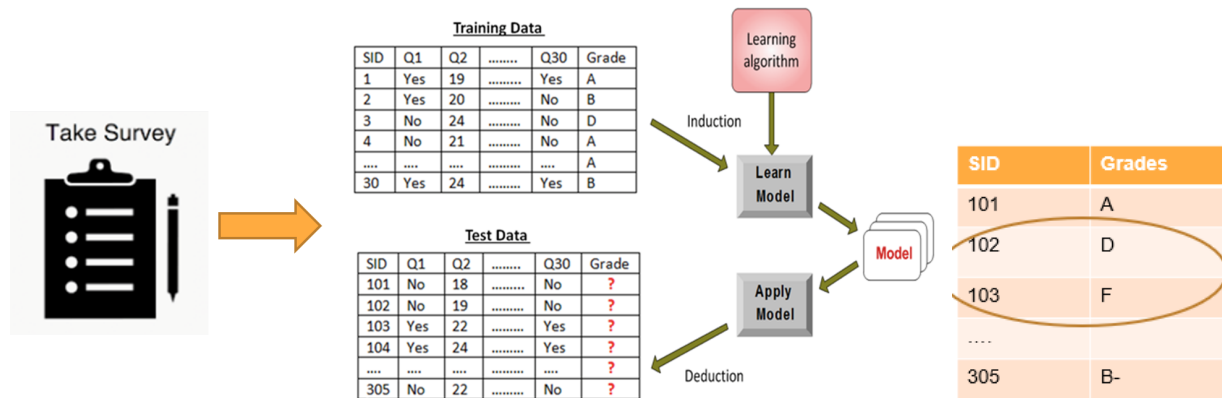

Fig 3: Pipeline showing methodology of predicting grades. 30 survey (Q1-Q30) questions (including general, behavioral and course related questions), which helped predict end of semester grades of students. Training data from last 5 years with the same survey questions were used.

CLASS ROLL	Final Exam Grade	Predicted Final Grade
	B+	B+
	A-	A-
	B+	B+
	D	D
	A	A-
	B+	B-
	B+	B+
	D	D
	B+	A-
	A	A
	A-	B+
	B	B-
	B	B
	B	B-
	B+	B+
	D	D
	A-	B
	B+	B+
	B+	B
	B+	A-
	A	A-
	A-	A-
	A-	A-

Table 2: Showing the predicted grades (after the first month into the semester) and actual grades (at the end of the semester) in Fall 2021 (of a section of 24 students), with 91% accuracy in predictions.

5.2 Results:

Table 3 shows the results of four semesters of using the ensemble classifier to predict grades of students in the CST department from 4 courses. We plan to expand it to other courses over the coming years. How the data was used and what prescriptions were offered to the students are discussed in the next section (section 6(a)).

Number of students	Semester	Accuracy % of grades (A-F) by the ensemble classifier
234	Fall 2020	87%
347	Spring 2021	84%
489	Fall 2021	91%
659	Spring 2022	88%

Table 3: Four semesters of using the predictive tool to predict grades of students based on prior data and ensemble classification.

6. Prescriptive tools:

Prescriptive data analytic tools are useful in academic departments for making decisions by combining descriptive and predictive analytics with optimization algorithms to generate recommendations for specific actions or decisions. We use prescriptive tools to do the following things at our department:

(a) Student outcomes and Retention rates: In the case of student performance, our prescriptive analytics model uses data on previous academic performance, attendance, and demographic information to make predictions about future grades (see the previous section on Predictive tools), and then use *optimization algorithms* to generate recommendations for targeted interventions or resources (mentoring, tutoring and seat allocations) that are most likely to improve student performance. This information can then be used to provide targeted support to students who are at risk of underperforming and help them to improve their grades.

After the student grades were predicted by the tool (see Table 2), The students whose grades were predicted to be lower than C, were (a) recommended tutoring and/or mentoring, (b) paired with those students who were predicted to get higher grades (B or higher), and (c) recommended seats in the front of the classroom. The system directly emails the students and the predicted grades were not revealed to the professor teaching the course. This is to avoid any implicit bias the professor may have towards the students

on seeing the predicted grades. The students would then avail the resources provided by the system.

Results: The retention rates and student academic performances are shown in Table 4 and Table 5. There was statistical significance in the performance of students (using Chi-square goodness of fit test).

(a)

	Total	Grades A-C		Grades A-D		Grades W	
		#	%	#	%	#	%
No-DA	285	197	69%	217	76%	28	10%
DA	185	148	80%	157	85%	5	3%

(b)

	Retention		
CS0	# retained	Total	%
No-DA	242	285	85%
DA	174	185	95%

Table 4: For an introductory CS0 course, (a) shows the improvement in student performance and (b) shows the one year retention rates.

Number of students	Semester used	One year freshmen retention rates (without Data Analytics)	One year freshmen retention rates (students who used Data Analytics)
578 (DA) / 1068 (non-DA)	Fall 2020	62.3%	69.7%
788 (DA) / 892 (non-DA)	Fall 2021	61.2%	70.8%

Table 5: Improvement in one year freshman retention rates at the entire CST department (before and after using data analytics). In row 1, we see that 578 students (DA) used the methodology described and 1068 students (non-DA) were not part of the data analytics treatment.

(b) Faculty: In the case of faculty performance, our prescriptive analytics model uses data on research productivity, grant funding, and teaching evaluations (see section on descriptive analytics) to make suggestions about future performance of tenure track faculty and then use *optimization algorithms* to generate recommendations for targeted support and resources that are most likely to improve faculty performance (for teaching - summer teaching workshops, for research - specific NSF workshops, and research collaboration workshops). This information can then be used to provide targeted support to faculty members who are at risk of underperforming and help them to improve their performance.

(c) Budget and Resource allocation: At the institution level we are currently using four prescriptive optimization tools: *Linear programming, Integer programming, Mixed Integer programming and Nonlinear programming* which prescribe cost-cutting based on the following four dimensions:

  (i) Staff workload: By analyzing data on staff workload and productivity, prescriptive analytics helps identify areas where staff members are overworked or underutilized. This can inform decisions about how to redistribute staff resources or where to invest in additional staff.

  (ii) Course enrollment: By analyzing data on course enrollment patterns, prescriptive analytics helps our CST department administrators make decisions about how to allocate teaching resources. For example, analytics has helped identify courses that are consistently oversubscribed and where additional teaching resources may be needed, or courses that are consistently undersubscribed and where resources can be reallocated.

  (iii) Facility usage: By analyzing data on facility usage, prescriptive analytics helps the CST department make decisions about how to allocate physical resources such as classrooms, labs, and equipment. For example, analytics can identify which facilities are being overused or underused, and where changes to scheduling or resource allocation may be needed.

Semester used	Cost cutting area / Planning	% of savings by implementing prescribed methods
Fall 2021		7.1%
Spring 2022	Facilities	8.6%
Fall 2022		7.9%

Table 6: Cost-savings by implementing recommendations of prescriptive tools

(iv) Grant funding: By analyzing data on grant funding, prescriptive analytics can prescribe departments on how to allocate resources to research projects. For example, analytics can identify which research projects are likely to be successful based on historical data and can inform decisions about how to allocate funding and resources to those projects.

Table 7 shows a summary (aggregated over 3 years) of all cost savings by implementing several prescriptive algorithms .

Semester used	Cost cutting area / Planning	% of savings by implementing prescribed methods
Fall 2022	Staff (Lab, College assistants, etc)	3%
Fall 2022	Faculty reduction	2%

Table 7: Estimation of cost cutting with some prescriptions from the prescriptive tool (using aggregated data from 3 years)

Discussion

In this paper we identified how analytics can help departments make useful decisions:
(a) Data-driven decision-making: Analytics enables academic departments to make decisions based on data rather than assumptions or intuition, which can lead to more objective and accurate decision-making.
(b) Increased efficiency: Analytics can help academic departments identify inefficiencies in their operations, which can lead to cost savings, improved resource allocation, and increased productivity.
(c) Improved student outcomes: Analytics can help academic departments identify patterns in student performance, such as areas where students are struggling or excelling, which can inform decisions about how to improve teaching and support services.
(d) Better resource allocation: Analytics can help academic departments allocate resources more effectively, such as determining which courses need more or fewer teaching resources, or which research projects are most likely to be successful.

However, there are certain cons of using data analytics in departments:
(a) Limitations of data: Analytics relies on the quality and quantity of data available. If the data is incomplete or inaccurate, it can lead to flawed decision-making.

(b) Cost: Implementing analytics tools and systems can be expensive, which may be a barrier for some academic departments, particularly those with limited resources.

(c) Resistance to change: Some faculty and staff may be resistant to using analytics to make decisions, particularly if they are not familiar with the technology or feel that it undermines their expertise.

(d) Ethical concerns: There may be ethical concerns around the use of student data, particularly in terms of privacy and data security.

Conclusion

In this paper, we have demonstrated three data-analytic strategies: predictive, descriptive and prescriptive and how they have improved student outcomes, intervened at-risk students, strategized cost cutting in the department, project actual outcomes and finally determining the effectiveness of our data-decisions. Our empirical results show that effective usage of data analytics with simple algorithms (k-means, random forest, decision trees, etc) can show considerable improvement at the department level (with retention rates), at the college level (with faculty teaching improvement) and at the university level (with cost-cuts).

References

[1] Kharade, B., & Wagh, K. (2016). Data Analytics in Educational Management System. International Journal of Computer Applications, 975, 8887.

[2] Nghe, N. T., Janecek, P., & Haddawy, P. (2007, October). A comparative analysis of techniques for predicting academic performance. In 2007 37th annual frontiers in education conference-global engineering: knowledge without borders, opportunities without passports (pp. T2G-7). IEEE.

[3] Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. Procedia Technology, 25, 326-332.

[4] Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. International Journal of Information Management, 50, 57-70.

[5] Shayakhmetova, B. K., Ayazhanov, S. S., Ten, T. L., Drozd, V. G., & Medeubaeva, A. M. (2021, April). Descriptive big data analytics in the field of education. In 2021 IEEE international conference on smart information systems and technologies (SIST) (pp. 1-5). IEEE.

[6] Jesenko, D., Brumen, M., Lukač, N., Žalik, B., & Mongus, D. (2018, March). Visualization and analytics tool for multi-dimensional data. In Proceedings of the 2018 International Conference on Big Data and Education (pp. 11-15).

[7] Hodgman, S. B. (2008). Predictive modeling & outcomes. Professional case management, 13(1), 19-23.

[8] Nafea, I. T. (2018). Machine learning in educational technology. Machine learning-advanced techniques and emerging applications, 175-183.

[9] Gupta, D. (2018). Applied analytics through case studies using Sas and R: implementing predictive models and machine learning techniques. Apress.

[10] Guan, J., Nunez, W., & Welsh, J. F. (2002). Institutional strategy and information support: the role of data warehousing in higher education. Campus-wide information systems, 19(5), 168-174.

[11] Aisch, G., Buchanan, L., Cox, A., & Quealy, K. (2017, January 18).. New York Times. Retrieved from https://www.nytimes.com/interactive/2017/01/18/upshot/some-colleges-have-more-students-from-the-top-1-percent-than-the-bottom-60.html

[12] Singh, C., Gopal, A., & Mishra, S. (2011, April). Extraction and analysis of faculty performance of management discipline from student feedback using clustering and association rule mining techniques. In 2011 3rd International Conference on Electronics Computer Technology (Vol. 4, pp. 94-96). IEEE.

[13] Davidson, I., & Satyanarayana, A. (2003, November). Speeding up k-means clustering by bootstrap averaging. In IEEE data mining workshop on clustering large data sets.

[14] Esterhuizen, H. D., Blignaut, S., & Ellis, S. (2013). Looking out and looking in: Exploring a case of faculty perceptions during e-learning staff development. International Review of Research in Open and Distributed Learning, 14(3), 59-80.

[15] Satyanarayana, A., Lansiquot, R., & Rosalia, C. (2019, October). Using prescriptive data analytics to reduce grading bias and foster student success. In 2019 IEEE Frontiers in Education Conference (FIE) (pp. 1-5). IEEE.