Research Statement

Manoj Thulasidas
Associate Professor of Computer Science (Education)

I. Background

With the recent changes in the school-wide research focus and objectives, it is time to take a fresh look at my research interests and perhaps pivot from pedagogical topics toward more core or theoretical themes. My current research interests and topics have been largely pedagogical in nature. I have had some success in forging collaborations and making impactful publications in that area, culminating in a Best Paper Award at TALE 2023 for *"Peer Learning in an Undergraduate Linear Algebra Course - A Social Network Analysis."* In the next phase, I plan to get into more advanced ideas in learning analytics as well as the field of quantum computing, which enjoys currency and popularity among computer scientists and physicists alike.

II. Research Areas

In the past, I focused on the four distinct research areas listed below. However, in light of the change in the research posture of the school, future efforts will fall mainly in the first category.

A. *Algorithms:* Drawing on my quantitative and scientific background, I have delved into creating foundational algorithms that illustrate the fundamentals of data science and statistical concepts. This area of research has led to the publication of a few conference papers that contribute to the advancement of teaching the basics of analytics and statistics with a focus on their relevance and applicability in algorithms.
   As a new area where I may be able to make an impact, I have started working with a colleague and a UG student on some applications and refinements of Quantum Computing. Although it is too early to pinpoint the output, I plan to expand my research efforts in this area in the coming years, with possible collaboration with other interested faculty with a background similar to mine.

B. *Teaching Tools:* Another aspect of my research endeavors centers around the development of teaching tools and techniques that enhance the efficacy and productivity of educators. These tools elevate the overall educational experience for both students and instructors, thereby fostering an environment of enhanced learning and engagement.

C. *Pedagogy:* Course design and classroom activities and their correlation with the student learning experience form the third aspect of my pedagogical research interests. This category of papers includes the analysis of student feedback and reflections with a view to understanding the impact of teaching practices.

D. *Popularization:* Presenting complex topics in a more approachable and comprehensible manner instills a greater appreciation for the subject matter and promote knowledge dissemination beyond the confines of academia. With this in mind, I have published articles, podcasts, videos and textbooks that serve as accessible resources for a broader audience.

The articles and other publications under each of these categories are summarized in Table I. Here are short summaries of the abstracts of these articles, grouped by the category to which they belong:

A. Algorithms

1) A Quality Metric for K-Means Clustering: This paper addresses the use of the K-Means algorithm in introductory data analytics courses. While K-Means is conceptually simple, it has drawbacks related to variable selection and determining the optimal number of clusters. The paper introduces a new quality metric called Standard Score Metric (SSM) based on the centroids' distribution, providing a mathematically defensible approach. The SSM is demonstrated for automatic variable selection and optimal cluster determination using both well-known datasets and locally collected real data.

2) Nearest Centroid: A Bridge between Statistics and Machine Learning: This paper introduces the Nearest Centroid (NC) algorithm as a pedagogical tool for machine-learning students to understand statistical thinking. NC combines key concepts from K-Means clustering and K Nearest Neighbors (k-NN) algorithms. The paper demonstrates the application of probability and statistics in machine learning and explains practical aspects of validation and performance measurement. The algorithm can be used in labs and reading assignments to reinforce students' understanding of applied statistics and its connection to machine learning.

3) A Quality Metric for K-Means Clustering Based on Centroid Locations: This paper tackles the problem of determining the right number of clusters in the K-Means clustering algorithm, which lacks a clear K selection methodology. It presents a novel metric, based on centroid locations and desired cluster properties. Developed in two stages, it considers the full covariance matrix and later extends to accommodate K-Means assumptions. Demonstrated on synthetic and real datasets, the metric's efficacy in identifying the optimal cluster count is showcased, along with comprehensive comparisons to existing quality indexes for automatic cluster determination.

4) A Recommendation on How to Teach K-Means in Introductory Analytics Courses: K-Means clustering is a widely taught unsupervised machine learning algorithm in data analytics courses for its simplicity and practicality. However, a challenge it encounters is determining the suitable number of clusters (K selection). The commonly used elbow method examines changes in quality metrics like the sum of squared errors (SSE). This paper surveys different metrics and methods for K selection, suggesting the Variance Ratio Criterion (VRC) as an appropriate metric to teach due to its desirable mathematical properties, enhancing students' understanding of the algorithm's statistical foundations. The paper also highlights key concepts addressed by the VRC approach and proposes related assignments, aimed at enhancing student understanding.

Table I
SUMMARY OF RESEARCH OUTPUT

Category	Title	Venue	Status
Algorithms	A Quality Metric for K-Means Clustering	ICNC-FSKD 2018	Published
Popularization	Quantum Computing is Here to Stay	MITB Thought Leadership Series	Published
Popularization	Making Sound Decisions Through Data Analytics	SMU Podcast	Published
Algorithms	Nearest Centroid: A Bridge between Statistics and Machine Learning	TALE 2020	Published
Teaching Tools	Secure Answer Book and Automatic Grading	TALE 2020	Published
Popularization	Data Analytics: A Tool for Business Revival	IEEE Computer Society (2020)	Published
Popularization	Data Analytics: Bridging Technology and Business	Keynote address at IPECS 2020	Published
Teaching Tools	Statistical Moderation: A Case Study in Grading on a Curve	TALE 2021	Published
Popularization	Linear Algebra for Computer Science	Textbook for CS103 (2021)	Published
Algorithms	A Quality Metric for K-Means Clustering Based on Centroid Locations	ADMA 2022	Published
Algorithms	A Recommendation on How to Teach K-Means in Introductory Analytics Courses	TALE 2022	Published
Teaching Tools	Cold Calls to Enhance Class Participation and Student Engagement	TALE 2022	Published
Pedagogy	Peer Learning in an Undergraduate Linear Algebra Course - A Social Network Analysis	TALE 2023	**Best Paper Award**
Teaching Tools	The Grader: A Grading Assistant for Lab Tests in Programming Courses	EDUCON 2024	Published
Pedagogy	Flipped Classroom for Linear Algebra at Undergraduate Level	EDUCON 2024	Published

B. Teaching Tools

1) Secure Answer Book and Automatic Grading: To address the increasing demand for online assessments, I created a secure answer book and an automated grading tool for our spreadsheet modeling course. These innovations were applied to approximately 160 students. This paper outlines the design, implementation, and techniques used to enhance the security of the answer book and improve grading efficiency, accuracy, and consistency. Furthermore, it shares the valuable insights gained from both the instructor and student perspectives. While the answer book and grading tool were tailored for the specific course, their design principles are adaptable to other courses and electronically submitted assignments.

2) Statistical Moderation: A Case Study in Grading on a Curve: The practice of "grading on a curve" is often viewed negatively due to concerns about its influence on final grades based on cohort strength. However, achieving absolute uniformity in assessment across different years, instructors, and settings can be challenging. As a result, grading on a curve may be seen as a necessary approach to standardize final scores and enable fair comparisons between student cohorts. This paper presents grading on a curve as a form of statistical moderation, along with a moderation tool designed for grade standardization. It discusses the strategy and principles guiding its implementation and share our experiences with this approach.

3) Cold Calls to Enhance Class Participation and Student Engagement [Second Author: Aldy Gunawan]: This study explores the impact of cold calls on student engagement in the classroom. An automated system for unbiased, randomized cold calling is introduced, where students are given time to reflect before selecting a specific student to respond. The study finds a statistically significant increase in class participation and student engagement when the cold-calling system is used. The results suggest that cold calls positively influence student involvement in class activities. Future plans and study limitations are also discussed.

4) The Grader: A Grading Assistant for Lab Tests in Programming Courses [Second Author: David Lo]: This article introduces the Grader, a grading assistant application designed for a Web Application Development course. The Grader efficiently manages lab tests, including file management, rubric application, and auto-grading with test cases. It also detects cheating attempts, making it valuable for programming courses with lab tests. Developed in the same programming environment as the course, the Grader serves as a pedagogical tool, applying classroom techniques in a real-world context. The article covers the Grader's features, deployment in the course, and our experiences using this innovative tool.

C. Pedagogy

1) Peer Learning in an Undergraduate Linear Algebra Course - A Social Network Analysis [Second Author: Kyong Jin Shim. Third Author: Jonathan Teo (Student)]: This study utilizes Social Network Analysis (SNA) to investigate peer learning behaviors among undergraduate Linear Algebra students. Through SNA, the study examines interaction patterns, information flow, and collaboration within the classroom. The analysis identifies the prevalence and development of peer learning and its impact on students' academic performance. Additionally, it reveals the characteristics of students engaging in peer helping and the formation of small learning communities. The study's insights can be valuable for educators seeking to enhance peer learning and educational practices in Linear Algebra, contributing to educational reform efforts.

2) Flipped Classroom for Linear Algebra at Undergraduate Level: This article details the development of an undergraduate Linear Algebra course tailored to emphasize its relevance and applications in Computer Science. Through a three-year transformation, the course evolved from traditional direct instruction to a flipped-classroom design, resulting in positive student learning outcomes. The article discusses the course design philosophy, syllabus, learning objectives, and the integration of student feedback to shape the curriculum. Valuable insights are shared, serving as best practices for instructors delivering successful Linear Algebra courses to undergraduate Computer Science students. The study highlights the effectiveness of the flipped-classroom approach in engaging students and enhancing

their understanding of mathematical concepts, utilizing innovative teaching methods, and emphasizing practical applications in Computer Science education.

D. Popularization

In addition to my peer-reviewed publications, I have authored several other works, including a significant textbook titled "Linear Algebra for Computer Science" designed specifically for CS103. The book is accompanied by its own website, LA4CS.com, and a collection of videos created for the flipped-classroom pedagogy employed in the course. These resources have been published with minimal restrictions, making them accessible for anyone to utilize and benefit from.

III. Future Plans

Going forward, I plan to continue my research and publication efforts primarily in the four categories as above. Leveraging my background in physics, I will also look into quantum computing and how it applies to optimization problems and in finance. Here are some of the projects already in the pipeline or planned, in the four categories listed on the first page of this statement:

A. Algorithms

1) A Recursive K-Means Algorithm: One drawback of K-Means algorithm lies in its assumption of approximate uniformity in cluster membership and variance. If cluster variances differ significantly, K-Means clustering may fail. Normalizing data using the variance of corresponding clusters could mitigate this issue. However, the lack of cluster affiliation *before* clustering poses a recursive problem. This research project aims to break the recursion using the proposed algorithm.

2) Topic Painting: Can we attach topics to sentences in the documents?: This study proposes a novel approach to associate topics with sentences within documents using Latent Dirichlet Allocation (LDA). We will pre-process words for each topic to create a sparse vector in the vocabulary space. Then, we will tokenize the documents, identify dominant topics, and assign topics to sentences based on their distances from the dominant topics. We also plan to conduct sentiment analysis on each cluster, representing a dominant topic, using a review dataset from TripAdvisor and movie reviews. If successful, the results will enhance document understanding and advance information retrieval and text-based applications.

3) SVD for Clustering: In this research project, we explore the use of matrix factorization, similar to LSA in Text mining, for improving the efficiency of cluster assignment to observations in M-Means. By considering the left singular vectors (u_i) as features and the right singular vectors (v_i) as linear combinations of observations, we aim to select the first M singular vectors as centroids and use a k-NN kind of algorithm to assign the remaining vectors. We also investigate if this approach has been previously explored and consider the potential of using u_i for feature reduction. Additionally, we explore the possibility of using matrix factorization to serve as seeds for the K-Means algorithm. This topic is inspired by CS103, and will likely include its students participating in the project.

B. Teaching Tools

1) Peer Evaluation Tool: I plan to develop a Peer Evaluation Tool specifically designed for assessing projects, both within and across project teams. The tool aims to provide a fair and efficient process for students to evaluate their peers' contributions and teamwork during group projects, as well as how they evaluate other projects. By using this tool, students will be able to provide feedback on their team members' performance, collaboration skills, and individual contributions to the project. The Peer Evaluation Tool will serve as a resource for instructors to assess the overall effectiveness and fairness of group projects and ensure that each student's efforts are appropriately recognized and rewarded. The tool's implementation is expected to be an effective aid in promoting teamwork and accountability in project-based learning settings.

C. Pedagogy

1) Tracking of Learning Objectives and Competencies using Online Quizzes: One advantage of using online quizzes for continuous or summative assessments is the wealth of data it generates. This paper will adopt a case-study style approach to analyze the statistics from continuous assessments and final exam quizzes on our eLearn platform for two courses, COR1305 and CS103. By incorporating learning objectives and competencies, the study aims to quantify the coverage of the assessments and assess the efficacy of teaching based on student performance. The insights gained from this analysis will inform improvements in the course's teaching approach, benefiting future students.

2) Insights from Student Reflections and Performance: This project aims to correlate student grades with their reflective feedback and continuous assessment data, using text mining techniques for sentiment analysis and topic modeling. By analyzing the qualitative reflections alongside quantitative ratings, we seek to improve course delivery and identify areas of enhancement. The study will employ continuous assessment to gather data throughout the academic year, comparing the learning patterns of Master of Information Technology in Business (MITB) and undergraduate (UG) students. Additionally, we will investigate the differences in learning outcomes between traditional and flipped classroom approaches. The project's findings will facilitate evidence-based decision-making for optimizing the learning journey in diverse academic programs.

D. Popularization

I have a keen interest in writing about various aspects of technology, computer science and their philosophical backdrop. These articles or videos, which aim to share my thoughts and insights, will be published in suitable venues. Some of the topics I plan to explore include:

- The intelligence displayed by ChatGPT in its responses: What does it reveal about knowledge, intelligence, wisdom, and related concepts?
- The fascinating world of quantum computing: Unraveling the magic behind its awesome power.
- The significance of learning linear algebra in computer science: Understanding its role and relevance in the field.

IV. Closing Remarks

To summarize, I am pleased to witness the steady growth of my research output over the past four years, although it took a few years to gain momentum. As an education-track faculty, I have found a meaningful niche where my research contributions have made a positive impact and have spanned diverse areas. As I look to the future, my goal is to continue elevating both the quantity and quality of my research endeavors, seeking new opportunities to forge collaborations, inspire students and make a valuable and lasting contribution to the field.