THE ATHENS INSTITUTE FOR EDUCATION AND RESEARCH

Abstract Book

9th Annual International Conference on Industrial, Systems and Design Engineering
21-24 June 2021, Athens, Greece

Edited by
Gregory T. Papanikos

2021

Abstracts
9th Annual International Conference on Industrial, Systems and Design Engineering

21-24 June 2021, Athens, Greece

Edited by Gregory T. Papanikos

First published in Athens, Greece by the Athens Institute for Education and Research.
ISBN: 978-960-598-421-2

8 Valaoritou Street
Kolonaki, 10671 Athens, Greece
www.atiner.gr

TABLE OF CONTENTS

(In Alphabetical Order by Author's Family Name)

Preface

This book includes the abstracts of all the papers presented at the *9th Annual International Conference on Industrial, Systems and Design Engineering* (21-24 June 2021), organized by the Athens Institute for Education and Research (ATINER).

A full conference program can be found before the relevant abstracts. In accordance with ATINER's Publication Policy, the papers presented during this conference will be considered for inclusion in one of ATINER's many publications.

The purpose of this abstract book is to provide members of ATINER and other academics around the world with a resource through which to discover colleagues and additional research relevant to their own work. This purpose is in congruence with the overall mission of the association. ATINER was established in 1995 as an independent academic organization with the mission to become a forum where academics and researchers from all over the world could meet to exchange ideas on their research and consider the future developments of their fields of study.

It is our hope that through ATINER's conferences and publications, Athens will become a place where academics and researchers from all over the world regularly meet to discuss the developments of their discipline and present their work. Since 1995, ATINER has organized more than 400 international conferences and has published nearly 200 books. Academically, the institute is organized into 6 divisions and 37 units. Each unit organizes at least one annual conference and undertakes various small and large research projects.

For each of these events, the involvement of multiple parties is crucial. I would like to thank all the participants, the members of the organizing and academic committees, and most importantly the administration staff of ATINER for putting this conference and its subsequent publications together. Specific individuals are listed on the following page.

Gregory T. Papanikos
President

9th Annual International Conference on Industrial, Systems and Design Engineering, 21-24 June 2021, Athens, Greece

Organizing & Scientific Committee

All ATINER's conferences are organized by the Academic Council. This conference has been organized with the assistance of the following academic members of ATINER, who contributed by reviewing the submitted abstracts and papers.

1. Gregory T. Papanikos, President, ATINER & Honorary Professor, University of Stirling, U.K.
2. Theodore Trafalis, Director, Engineering & Architecture Division, ATINER, Professor of Industrial & Systems Engineering and Director, Optimization & Intelligent Systems Laboratory, The University of Oklahoma, USA.
3. Timothy M. Young, Director, Center for Data Science (CDS) & Professor and Graduate Director, The University of Tennessee, USA.
4. Dimitrios Goulias, Head, Civil Engineering Unit, ATINER and Associate Professor & Director of Undergraduate Studies Civil & Environmental Engineering Department, University of Maryland, USA.

FINAL CONFERENCE PROGRAM
9th Annual International Conference on Industrial, Systems and Design Engineering, 21-24 June 2021, Athens, Greece

PROGRAM

Monday 21 June 2021

09.30-10.00
Registration

10.00-10.30
Opening and Welcoming Remarks:

- **Gregory T. Papanikos**, President, ATINER.
- **Timothy M. Young**, Director, Center for Data Science (CDS) & Professor and Graduate Director, The University of Tennessee, USA.
- **Theodore Trafalis**, Director, Engineering & Architecture Division, ATINER, Professor of Industrial & Systems Engineering and Director, Optimization & Intelligent Systems Laboratory, The University of Oklahoma, USA.
- **Dimitrios Goulias**, Head, Civil Engineering Unit, ATINER and Associate Professor & Director of Undergraduate Studies Civil & Environmental Engineering Department, University of Maryland, USA.

10.30-11.15
Tim Young, Director, Center for Data Science (CDS) & Professor, University of Tennessee at Knoxville, USA.
Debdatta Saha, Professor Faculty of Economics, South Asian University, India.
Title: *Predicting and Explaining Firm Size in India with Machine Learning.*

11.15-11.30 Break

11.30-12.15
Moshe Eben Chaime, Professor, Ben Gurion University of the Negev, Israel.
Title: *Project Planning and Control from Engineering Perspective.*

12.15-12.30 Break

12.30-13.15
Tim Young, Director, Center for Data Science (CDS) & Professor, University of Tennessee at Knoxville, USA.
Chloe Ericksen, Graduate Research Assistant, University of Tennessee, USA.
Title: *Developing an AI Platform For Continuous Manufacturing Systems.*

13.15-13.30 Break

13:30-14:15
Iralda Xhaferaj, Lecturer, Polytechnic University of Tirana, Albania.
Neritan Shkodrani, Professor, Polytechnic University of Tirana, Albania.
Title: *Road Safety Impacts of Sight Distance Criteria According to Albanian Code.*

14:15-14:45 Lunch

14:45-15:30
Amer Alnuaimi, Adjunct Professor, San Diego State University, USA.
Title: *Evaluation of Multi-Story Buildings According to the IRAQI Seismic Code Requirements.*

Tuesday 22 June 2021

09:30-10:15
Peter Sudermann, Head of Official Test Facility for Building Materials Koblenz, University of Applied Science, Germany.
Manfred Breitbach, Professor, Koblenz University of Applied Science, Germany.
Title: *Concrete Structures for Drinking Water Reservoirs.*

10:15-10:30 Break

10:30-11:15
Andrius Baltrusaitis, PhD Student, Vilnius Gediminas Technical University, Lithuania.
Title: *Comparison of Asphalt Pavement Compaction Quality Determination Using Ground Penetrating Radar and other Traditional and Non-Destructive Testing Methods.*

11:15-11:30 Break

11:30-12:15
Raphael Riva, Post-Doc Researcher, University of Liege, Belgium.
Title: *Low Chemistry as Innovative Approach for Sustainable Formulations.*

12:15-12:30 Break

12:30-13:15
Sheikh Shahriar Ahmed, Graduate Research Assistant The State University of New York, USA.
Sarvani Sonduru Pantangi, Research Associate Michigan State University, USA.
Grigorios Fountas, Lecturer Edinburgh Napier University, UK.
Kevin Majka, Senior Research Scientist Public Safety & Transportation Group, USA & Panagiotis Anastasopoulos, Associate Professor The State University of New York, USA.
Title: *Improving Pedestrian Safety Using High Visibility Crosswalks: A Correlated Grouped Random Parameter Approach with Heterogeneity in Means.*

13:15-14:00 Lunch

14:00-14:45
Levent Yilmaz, Professor, Auburn University, USA.
Title: *Simulation Credibility Assessment via Dynamic Introspective Cognitive Trust Models.*

14:45-15:00 Break

15:00-15:45
Kenneth Lee, Professor, Western New England University, USA.
Title: *Viscosity Variations of Each Liquid in a Three-Component Isopropanol-Hexane-Water System.*

15:45-16:00 Break

16:00-16:45
Dimitrios Goulias, Head, Civil Engineering Unit, ATINER and Associate Professor & Director of Undergraduate Studies Civil & Environmental Engineering Department, University of Maryland, USA.
Title: *Incorporating NDT in Quality Assurance of Concrete & Performance of Bridge Decks with ML.*

Sheikh Shahriar Ahmed
Graduate Research Assistant, The State University of New York, USA
Sarvani Sonduru Pantangi
Research Associate, Michigan State University, USA
Grigorios Fountas
Lecturer, Edinburgh Napier University, UK
Kevin Majka
Senior Research Scientist, Public Safety & Transportation Group, USA
&
Panagiotis Anastasopoulos
Associate Professor, The State University of New York, USA

Improving Pedestrian Safety Using High Visibility Crosswalks: A Correlated Grouped Random Parameter Approach with Heterogeneity in Means

This study seeks to evaluate the effectiveness of high-visibility crosswalks (HVCs) in terms of modifying driver behavior and improving pedestrian safety. For this purpose, the Second Strategic Highway Research Program (SHRP2) Naturalistic Driving Study (NDS) data are used, including information about SHRP2 participants' traversals before, during and after the installation of HVCs in various locations across the US. In this context, various aspects of driving behavior are analyzed for HVCs located at different positions on the block (mid-block vs. end-of-block) and different HVC marking designs (continental, bar-pair, and ladder). As no pedestrian-vehicle crashes or conflicts were identified from the forward-facing videos of traversals and the time series included in the SHRP2 NDS data, crash surrogate measures (i.e., speed, acceleration, throttle pedal actuation, and brake application) were employed to identify and analyze modifications in driving behavior at or near the HVCs.

To investigate the effect of HVCs on driving behavior, continuous and discrete outcome statistical models of the crash surrogate measures were estimated. The correlated random parameters statistical modeling framework was employed to account for a number of misspecification issues (panel effects arising from multiple traversals undertaken by each participant, the effect of unobserved characteristics, as well as the correlation among the unobserved characteristics as captured by the random parameters). Apart from the effect of HVCs, several other factors were explored and controlled for (e.g., vehicle and driver characteristics, roadside environment, and weather conditions) throughout the model estimation process. The results of the analysis

show that the HVCs have the potential to improve pedestrian safety, as well as positively modify driving behavior in the proximity of the crosswalks. Specifically, the HVC type (ladder or bar-pair) and in-block location (mid-block, or end-of-the-block) were found to improve safety by affecting the vehicles' speed, acceleration, throttle pedal actuation, and brake application. Finally, various trip-, traffic-, and driver-specific characteristics were also found to affect driving behavior at HVC locations.

Amer Alnuaimi
Adjunct Professor, San Diego State University, USA
&
Fouad Mohammad
Senior Lecturer, Nottingham Trent University, UK

Evaluation of Multi-Story Buildings According to the IRAQI Seismic Code Requirements

The aim of this study is to assess the performance objectives defined in the Iraqi Seismic Code (ISC) in order to make a realistic evaluation related to Performance Based Design (PBD) of multi-story reinforced concrete buildings and also to compare and evaluate structural response demands obtained from nonlinear static analysis procedures according to two versions of the displacement coefficient method (DCM) which are recommended in FEMA 356 and FEMA 440. Two groups of three-dimensional RC buildings with different heights, designed according to Iraqi Building Code Requirements for Reinforced Concrete (IBC), are investigated. Pushover analyses are carried out to determine the nonlinear behavior of the buildings under three different seismic hazard levels, for two Iraqi seismic zones, of earthquake loads. In order to determine performance levels of the buildings, maximum inter-story drift demands and plasticizing sequence are determined and compared with the related limits using the DCM recommended in FEMA 356 and FEMA 440. From the results of this research, it can be concluded that RC buildings designed according to the Iraqi codes sufficiently provide the performance objectives stipulated in the ISC. Comparing structural response quantities obtained from the two versions of DCM, effects on performance evaluations of the buildings are investigated comparatively, as well.

Andrius Baltrusaitis
PhD Student, Vilnius Gediminas Technical University, Lithuania

Comparison of Asphalt Pavement Compaction Quality Determination Using Ground Penetrating Radar and other Traditional and Non-Destructive Testing Methods

Asphalt pavement layers' compaction, expressed by air voids ratio between field and laboratory bulk density, assurance is one of main criteria of the durability of asphalt road pavement. Those, destructive measures should be applied and a lot of asphalt samples should be taken in field seeking to determine representative compaction level of constructed pavement. With the fast development of technologies new methods should be evaluated for fast, nondestructive and accurate asphalt bulk density air voids determination in field. Ground Penetrating Radar, can be such technology, allowing us to determine the qualitative characteristics of asphalt pavement across the entire length of the road without causing the damage to the road surface.

In order to apply this innovative method in Lithuania research was conducted in 2019 on a newly built 4-km long road section of highway A14 Vilnius–Utena located near Vilnius, Lithuania. The tests were performed on base and binder layers composed of AC 22 PS and AC 16 AS hot asphalt mixtures. Ground penetrating radar (GPR) measurements were conducted on each asphalt layer after they were constructed. 50 test points on each layer were selected and layer density, air void content and compaction were determined using two nondestructive technologies (NDT) Troxler, PQI and by drilling cores and testing them in the laboratory. To estimate asphalt layer density and air void content, GPR data was analyzed using different existing mathematical models. To justify the reliability of the GPR method, results were checked by comparing with other NDT and results measured directly on cores taken from the different asphalt layers.

Moshe Eben Chaime
Professor, Ben Gurion University of the Negev, Israel

Project Planning and Control from Engineering Perspective

The critical path method (CPM) has been a major break-through in modern management as it facilitates project management, thereby enabling effective and efficient management of gigantic projects almost independent of their size. Nevertheless, there are many issues that the CPM does not solve. Project management involves planning and control, among other tasks. Common to both, is the notion of criticality. What is critical and is this a fix or dynamic attribute? While planning is a mandatory task, it is relatively short compared to project control, which needs to escorts the project throughout its execution and maybe even after. Projects are aimed at the accomplishment of predefined goals within budget and on time. These three dimensions: scope, time and costs, are interrelated, but in project control there is a separation between issues; e.g., scope and quality, which pertain to goal accomplishment, and time and cost control. With regard to time and cost, the need became apparent "to decouple schedule (time) and cost dimensions" (Khamooshi and Golafshani, 2014). While this is correct for many, maybe even most elements, there are cost components which are time dependent. There are activities which have no measurable output but have to be conducted for certain time intervals. For each activity of this type, there is an event which triggers its performance and another event that brings it to termination. The cost of such an activity is proportional to its actual duration. Apparently, as comes out from the literature, this feature makes these cost components problematic for project control, but there seems to be a simple solution. In this talk, criticality issues, control of time dependent cost components and other issues will be discussed and examined.

Dimitrios Goulias
Head, Civil Engineering Unit, ATINER and Associate Professor & Director of Undergraduate Studies Civil & Environmental Engineering Department, University of Maryland, USA

Incorporating NDT in Quality Assurance of Concrete & Performance of Bridge Decks with ML

Non-destructive techniques (NDTs) employ non-invasive methods to evaluate the condition of materials and structural members. Other advantages of NDT include, in many cases, shorter testing time, lower inspection cost and higher inspection rates of products comparable to destructive testing, availability of real time results as production goes on, and minimizing bias of subjective visual inspections. Non-destructive techniques are able to evaluate the cover depth, concrete properties, internal flaws, and detect the reinforcement location and characteristics. Previous studies have explored to some extend specific NDTs in terms of accuracy and precision. However, in order to define NDT based acceptance for QA/QC purposes, further assessment is often needed in order to identify quality acceptance thresholds for concrete properties. One of the objectives of this study was to investigate and assess how specific non-destructive testing methods (NDTs) can be incorporated in the QA process of concrete and assess their ability to provide enough accuracy for the quality assurance of concrete and construction features of selected concrete members (concrete slabs, reinforced concrete members, bridge decks). Another objective of this study was predicting the performance condition of concrete bridge decks, based on national bridge inventory (NBI). By predicting the condition of bridge decks, the frequency of inspection, which is labor intensive and costly, can be reduced. For the former purpose, the experimental plan of the study considered a variety of different experiments designed to assess the response of the selected NDTs. Multiple NDTs were employed to test concrete at different ages and/or compare the response on flawed concrete elements in relation to sound concrete. For the latter purpose, predicting the future condition of bridge deck, two different machine learning methods, long-short term memory (LSTM) and convolutional neural network (CNN) were employed to classify the sequences of bridge condition ratings.

Kenneth Lee
Professor, Western New England University, USA
Ying Wang
Lecturer, National Engineering Laboratory for Coalmine Backfilling Mining, Shandong University of Science and Technology, China
&
Huigong Niu
Graduate Student, Department of Geology, Shandong University of Science and Technology, China

Viscosity Variations of Each Liquid in a Three-Component Isopropanol-Hexane-Water System

This experimental study investigates the variations of dynamic viscosity of each liquid phase in a two-phase, three-component system comprising isopropanol, hexane, and water. When combining a cosolvent, nonaqueous phase liquid, and water, the resulting mixture can either form a single-phase solution or a two-phase liquid-liquid system, depending on the composition of the three components. In the two-phase region, an aqueous phase and a nonaqueous phase are formed. The experiment starts by mixing a fixed ratio of isopropanol and hexane in the absence of water, which results in a single-phase mixture. The viscosity and the density values of the solution are then determined. Next, a small amount of water is added to the isopropanol-hexane solution, and the viscosity and density values are determined for the new mixture. The single-phase experiment repeats with increasing water content until the mixture partitions into a two-phase liquid-liquid system. The viscosity and the density values are then determined for each liquid phase. The two-phase experiment repeats by adding an incremental amount of water until the total water content is near 100%. A commercially available falling-ball viscometer is modified to determine the viscosity of each liquid phase within the two-phase region. Results show varying viscosity differences between the two liquid phases at different total water contents. Furthermore, the observed aqueous phase viscosity can be higher than the range of values set between the three individual components.

Raphael Riva
Post-Doc Researcher, University of Liege, Belgium

Low Chemistry as Innovative Approach for Sustainable Formulations

IN FLOW is an innovative cross-border R&D project involving 4 European cutting-edge partners in the fields of biochemical product formulation, characterization and engineering. The main R&D challenge of IN FLOW is to introduce new formulation technologies that allow fast and cheap degradable packaging of e.g. drugs in pills or creams. Our applied methodologies will allow companies to create novel high potential products for e.g. health- and personal care industry. The knowledge and technology that will be developed in IN FLOW provides public and private actors full advantages to fasten market introduction of their products. Our IN FLOW technology makes use of sophisticated in-house designed devices and know-how to create novel product formulations of (bio)degradable carriers loaded with ingredients with high efficiency. These carriers can be used as compounds in end-products like pills with direct applicability in healthcare, nutraceutical, cosmetic and pharmaceutical industries.

Precisely, the IN FLOW project focused mainly on the preparation of drug loaded microparticles made of degradable (co)polymers already successfully applied for biomedical applications, such as polycaprolactone or polylactide but also made of an emerging class of degradable and biocompatible polymers, namely polyphosphoesters. The selected microfluidics technology presents undeniable advantages for the formulation of drug loaded microparticles such as (i) synthesis in continuous flow avoiding possible fluctuation in composition between different production batches, (ii) very low polydispersity in size of the dug-loaded microparticles, (iii) high encapsulation efficiency, (iv) relatively low energy consumption, …

Via a market-driven approach, defined by private end-users (SMEs) through the open technology platform, IN FLOW will create innovative products and technologies. By combining unique Meuse-Rhin Euregional know-how, technologies and state-of-the-art infrastructures in an open innovation platform, IN FLOW proactively creates synergy for higher scale capacity for all actors.

Peter Sudermann
Head of Official Test Facility for Building Materials Koblenz, University of Applied Science, Germany
&
Manfred Breitbach
Professor, Koblenz University of Applied Science, Germany

Concrete Structures for Drinking Water Reservoirs

The constantly supply with healthy drinking water is one of the most important tasks for the World Health Organisation (WHO) and the European Commission. Furthermore drinking water reservoirs are necessary e.g., for fire protection, in times with ecological crises or war activities and basically to prevent epidemics. The hygiene and functionality requirements on are about 20 times higher than for usual food industries, because the population use their drinking water in many cases over the lifetime. Considering these circumstances there are very special requirements to the construction principles of the reservoir itself, material limitations concerning starting materials as well as material design (concrete, polymers, metallic materials), building supplies, repair materials and material changes due to the chemical attack by the drinking water and thereby negative influence on the further water quality. There must be a modeling of hydraulic functions and steady fluctuation of the water, stagnant water must be prevented in order to reduce microbiological growth.

In this contribution the essential aspects for open up construction projects, planning, joint constructions, crack width limitation, construction executions, quality assurance, hygienic standards for materials … will be explained by practical examples with new buildings and existing old structures (up to 100 years old) north of the Alps.

Iralda Xhaferaj
Lecturer, Polytechnic University of Tirana, Albania
&
Neritan Shkodrani
Professor, Polytechnic University of Tirana, Albania

Road Safety Impacts of Sight Distance Criteria According to Albanian Code

Sight distance is an important criterion in highway geometric design for traffic safety to ensure that the driver can see any possible road hazard insufficient time to take action and avoid an accident. Many Albanian roads have been designed and constructed before 1991 based on standards in place as of the construction time. More recently, investments have been made for upgrading and expanding the transport infrastructure in the country. The importance of providing adequate sight distance for safe and efficient traffic is well recognized by researchers and included in most design manuals. Horizontal curves, crest vertical curves, and rural intersections are the common sight restrictions considered in highway design. Stopping sight distance (SSD) is the most important of the sight-distance considerations since sufficient SSD is required at any point along the roadway. Technical road standards used in Albania before 1991 were based upon outdated standards. Currently, road design standards have been changed receiving considerable European standards. This paper analyses the sight distance estimation criterion using the upgraded manual. For this purpose, road design examples have been studied based on sight design requirements and geometric data. The results obtained through calculation were compared according to upgraded standards to investigate road safety associated with highway geometric design.

Levent Yilmaz
Professor, Auburn University, USA

Simulation Credibility Assessment via Dynamic Introspective Cognitive Trust Models

The increasing reliance of modern systems engineering on computer simulation demands appropriate trust in simulation models for credible results. Because of its foundations in operations research, model credibility is conventionally viewed from the lens of numerical and transformational accuracy. As simulation models facilitate exploration of explanations for scientific phenomena and the discovery of robust strategies under uncertainty, instilling confidence in their behavior is becoming paramount across a broad range of application categories. Among these categories include predicting system behavior, inferring the causal mechanisms of complex systems by exploring possibilities, supporting robust decision-making under uncertainty, training in virtual environments, intelligent tutoring in education, controlling cyber-physical systems, and diagnosis of behavior when a model reports its state.

In this paper, credibility is construed as a perceived measure of believability. As a perceived quality, it has multiple dimensions that are concurrently monitored to generate an evolving judgment. To this end, we examine the key concepts and terms that relate to credibility and then analyze the extant literature to discern emergent conceptual patterns. Based on this analysis and the recognition of the inherent challenges associated with the complexity and diversity of scientific problems, we propose a generic framework for the assessment of model credibility from the cognitive perspective of trust. Factoring in the cognitive perspective and the contextual interpretation of the multi-faceted characterization of model credibility is facilitated by associating credibility with trust. Despite the co-occurrence of credibility and trust as keywords in an increasing number of publications, especially in computational modeling for social and behavioral sciences, the current view of credibility within the Modeling & Simulation community is still centered around the notion of accuracy and empirical fitness.

To mitigate this limitation, we characterize model credibility in terms of the attributes, properties, and dynamics of trust. Three dimensions of trust and associated concepts are used to provide a foundation for trust-driven assessment of credibility. Fundamental processes in the formation and evolution of trust are highlighted to facilitate maintaining run-time trust models that can serve as cognitive

agents to establish appropriate, calibrated trust. Based on the characteristics of the model users, alternative evaluation models are recommended to measure and report trust metrics. We conclude with recommendations for establishing appropriate trust in simulation models. These considerations highlight design principles, provide strategies for relating contextual information to model capabilities for context-sensitive trust measurement, and emphasize the significance of factoring into the influence of the organizational culture and its community of practice.

Tim Young
Director, Center for Data Science (CDS) & Professor, University of Tennessee at Knoxville, USA
&
Chloe Ericksen
Graduate Research Assistant, University of Tennessee, USA

Developing an AI Platform for Continuous Manufacturing Systems

The advent of artificial intelligence (AI) as a standard for successful business endeavors has resulted in a 'Fourth Industrial Revolution' known as 'Industrie 4.0'. Industrie 4.0 is where computers and automation come together through remote connectivity to computer systems equipped with machine learning algorithms. Successful implementation of AI platforms is paramount to competing successfully in manufacturing. This study used automated data fusion with a suite of predictive algorithms to develop an AI platform for a continuous manufacturing system for bio-composites. An automated data fusion system using Microsoft SQL® was developed that aligned process sensor data with the destructive test data of product attributes. Process sensor data were aligned in time to correspond to the correct time at which the material flowed past the sensors to align with the final product development stage. Variable preselection was used with a suite of algorithms, e.g., partial least squares regression, neural networks, regression trees, boosted trees, bootstrap forest, and Bayesian Additive Regression Trees (BART). The product attributes of tensile strength and modulus were predicted real-time from the AI system. BART models predicted the best for both tensile strength and modulus. For tensile strength, RMSEP% ranged from 8.96% to 10.82% for the three largest product types manufactured. For modulus, RMSEP% ranged from 10.26% to 12.82% for the same three product types. System to line speed set point, adhesive set point, weight set point, wax set point, and moisture set point from operations personnel were compared with the feedback loop from the AI system for the same set points using the real-time AI predictions of tensile strength and modulus. The AI system made on average approximately 60% less adjustments across all of the set points. Line speed set point and adhesive set point were adjusted least by the AI system feedback loop. The real-time predictions of tensile strength and modulus had a CV% of 18%; while actual CV% of tensile strength and modulus had a CV% of 37%. The implication is that an AI system would make less adjustments

to key set points, and produce a final product of strength metrics with less variation.

Tim Young
Director, Center for Data Science (CDS) & Professor, University of Tennessee at Knoxville, USA
&
Debdatta Saha
Professor, Faculty of Economics, South Asian University, India

Predicting and Explaining Firm Size in India with Machine Learning

The performance of a firm in industry has long received attention from economists, both from empirical and theoretical lines of investigation. The reasons for the appropriate measure of firm size has been a contentious issue for economists for a long time. The objectives of the study were: 1) quantify the extent to which a firm's size affects its performance; 2) identify the hierarchy of independent variables in explaining performance; 3) identify the key dimension(s) of performance and firm size. Net fixed assets and 'real sales' (*sales deflated by WPI of FP [Rs. in millions*]) were the dependent variables used with four machine learning algorithms and 238 economic independent variables. The machine learning methods were: 'Boosted Trees'; 'Random Forests'; 'Regression Trees'; and 'Bayesian Additive Regression Trees (BART)'. Ten-fold validation was used to identify the hierarchy of independent variables in explaining performance and measure key dimension(s) of performance across the four machine learning algorithms. 'Real raw material exports' explained the most significant amount of variation in the models and ranged from 57% to 82% in total sums of squares depending on the algorithm. Other significant performance indicators, but to a lesser extent, that acted as a hierarchy of interactions with 'real raw material exports' were: 'real selling distribution for exports'; 'fixed assets'; 'real compensation to employees'; and 'power and fuel expenditures'. BART was consistently the better performer of the four algorithms in the majority of the 10 validations, with R^2 ranging from 0.565 to 0.983 for the ten validations. Boosted trees were second in predictive performance in validation. The study provides unique insights in using machine learning methods to understand performance of firms in India. BART results seem promising for predictions over a broad array of applications, business, economics, etc.