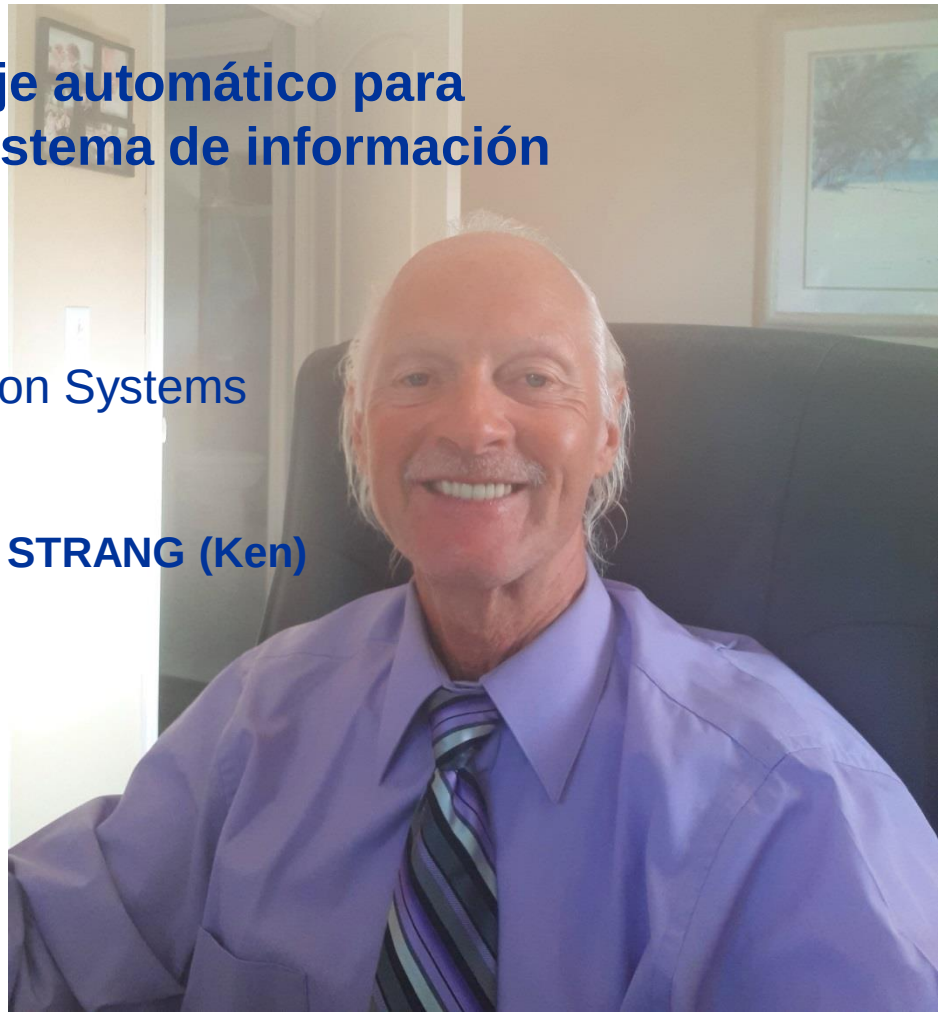
Presenter - Ken

Applying machine learning methods to understand unstructured information system big data

Aplicación de métodos de aprendizaje automático para comprender los grandes datos del sistema de información no estructurado



CISTI June 23, 2022 14:30 Room 6
Track: Organizational Models and Information Systems
Session Chair: Graca Azevedo

Principal investigator: **Professor Kenneth David STRANG (Ken)**
W3-Research, USA http://kennethstrang.com
professor@kennethstrang.com

Co-author: **Dr. Narasimha Rao VAJJHALA**
New York University at Tirana, Albania
narasimharaonarasimha@gmail.com
http://www.narasimharao.net

Motivation & Problem Rationale

- Researchers/Principal Investigator backgrounds
  - Ken: USA, Australia, others; Rao: Albania/EU, Nigeria, India
  - Multi-disciplinary: Computer science (IS, IT, ICT), management information systems, business; IEEE/ACM/PMI +++ members
  - 50+ years of combined business/IT experience, 300+ papers
- Rationale for current study
  - Historically approximately 50% of projects fail, inconclusive literature
    - Fail = failed to meet scope, budget, quality and or time mandates
      - (Borbath, Blessner & Olson, 2019, Eckerd & Snider, 2017, Strang, 2021)
      - "Overall 59% were successful, and 41% failed… classified 67.3% of the 2,692 projects … 12% effect size." (Strang, 2021, p. 30)
        - » Strang, K. D. (2021). Which organizational and individual factors predict success versus failure in procurement projects. International Journal of Information Technology Project Management, 12(3), 19-39. doi:10.4018/IJITPM.2021070102
  - A priori predictors of PM failure but only small 2%-12% effect sizes
  - Authors wanted to analyze large unstructured data from government
  - Mixed methods, programming, machine learning, then regression

Research Question & Methods

- Can machine learning (ML) find the failure causes by searching unstructured IS project big data?
- Unit of analysis: Government project unstructured big data
- A pragmatic mixed methods research design
- Critical analysis & PRIMSA literature review
- Structured programming (to fix up and extract big data)
- Random forest ML to identify likely important fields
  - Training – learning, effect sizes (27% in ML), ML accuracy
    - ROCa (AUC) 0.849 0.840
    - Sensitivity 0.810 0.526
    - Specificity 0.800 0.889
    - Precision 0.799 0.632
    - F1-measure 0.798 0.575

Results, Discussion, Conclusion

- Machine Learning Random Forest Feature Statistics

Feature	Mean accuracy decrease	Node purity increase
PM Experience (years)	0.037	0.063
Contract (outsourced) or in-house	0.007	0.058
Line of Business (special coding)	0.010	0.048
Remote work allowed for PM	0.010	0.032
ContractK (PM salary)	0.020	0.031
BudgetK (project level)	0.038	0.025
PM certified (APM, Prince, PMI..)	0.009	0.024

- **** Cut-off (mean decrease in accuracy positive, increase in node purity > 0.02)

- Logistic Regression on Project Cancelled (95% confidence)

Construct	Beta	Z	Wald	P
(Intercept)	-2.238	-4.023	16.185	< .001
PM Experience	-1.441	-9.434	88.996	< .001
PM ContractK	-0.113	-1.086	1.179	0.278
Project BudgetK	0.744	4.653	21.651	< .001
Outsourced/in-h (1)	0.869	11.629	135.224	< .001