Submitted to *Information Systems Research*
manuscript (Please, provide the manuscript number!)

Nascent Design Theory of Traffic Predictive Analytics Decisional Guidance

(Authors' names blinded for peer review)

To categorize and enable the generalization of the type of Information System proposed, we also articulate core prescriptions underpinning the Traffic Analytics Guidance Framework and Tool in the form of a nascent design theory.

Key words: Predictive Analytics; Decision Support; Decisional Guidance; Transportation Systems; IS Design Theories

---

Gregor and Jones (2007) describe theories for design and action as a class of highly influential theories in Information Systems (IS) focusing on the design, implementation, and use of artifacts underpinned by Information Technology (IT). According to Gregor and Jones (2007), "an IS design theory shows the principles inherent in the design of an IS artifact that accomplishes some end, based on knowledge of both IT and human behavior". IS design theories can underpin products or methods and provide a foundation for IS artifact construction. Artifacts involving ICTs use by humans distinguish IS design theories from design theories in disciplines such as architecture, medicine, and management (Gregor and Jones 2004). According to Simon (1981), design theories are theories of procedural rationality. Besides, design theories are prescriptive, dealing with goals, handling contingencies, and showing how artifacts that embody laws of interaction from natural and social sciences are put into practical use by embedding them into an IS design theory (Walls et al. 1992). A systematic theory specification approach is required for articulating design theories. The literature offers several templates for this task (Gregor and Jones 2007, Simon 1981, Walls et al. 1992, Kasper 1996). The template by (Gregor and Jones 2007) was chosen since it integrates previous work on design theory specification and provides comprehensive coverage of design theory dimensions.

Design Theory Specification

Purpose and Scope	TAG-F aims to provide systemic guidance to traffic data analysts in the Traffic Predictive Analytics (TPA) process lifecycle, through the structured characterization of the analytical problem space and the subsequent decision support via the recommendation of alternative predictive models to help analysts tackle complexity and find an effective Predictive Analytics Model for a traffic problem scenario. TAG-F guides traffic predictive analytics decision support by automating the identification and recommendation of predictive models and the ranking of available models/algorithms suitable for a given TPA problem/scenario. TAG-F is appropriate for supporting analytics stakeholders involved in traffic planning, control, and simulation. TAG-F is the first comprehensive analytical guidance framework that supports stakeholders in choosing a PAM for a traffic analytics problem scenario.

Constructs	Guidance in TPA is characterized as a function of a knowledge gap and a set of inputs, resulting in a desired output, either an answer or a means: Guidance (*Input*, *Analytical Requirements*) $\rightarrow$ *Output* *Input:* - Problem Specification (**PS**) - Domain Knowledge (**DN**) - Traffic data analyst's TPA expertise (**TPA_E**) [determines the degree of guidance needed via answers to system prompts] *Analytical Requirements* Data context (**DC**) - Prediction horizon - Data quality - Data set size - Analysis level Data collection method (**DCM**) - Sensor Device Type, Measurement Quality, Functionality. Predictive analytic/modeling method (**PAM**) - Literature-driven - Live project reports/results *Output* -PAM recommendation

Principles of Form and Function	**P1:** The framework should guide users (traffic analysts/traffic data analysts) by capturing users' input on features representing a given traffic analytics problem/scenario and ranking available models/algorithms suitable for a given TPA problem/scenario. **P2:** The framework should explore the relationships between the dimensions (DC, DCM, and PAM) to underpin recommendations of a PAM for a problem scenario. **P3:** The relationships between the dimensions (DC, DCM, and PAM) should represent mappings between features of a particular problem/scenario and a viable PAM selection for the problem/scenario. **P4:** The framework should create a meta-learning model for PAM selection by applying a combination of PAM knowledge extraction – derived using literature-based discovery (LBD) (Henry and McInnes 2017), and an instance-based learning inference model based on published traffic data science project results. **P5:** The concepts stated in (Constructs) should define the features of the meta-learning model with the meta-knowledge encoded in the model representing best practices derived from literature and extant instances of data science projects that encode the guidance provided. **P6:** The framework should provide an open and extensible architecture (depicted in Figure 4) where new PAMs are added to the corpus of literature and new data science projects in the traffic domain with published results feeding into the the feedback loop of the framework supporting more practical guidance.

Artifact mutability	The guidance provided by the framework should evolve due to the built-in feedback loop that captures: - New PAMs incorporated into the corpus of literature. - New traffic analytics project outcomes extending the corpus of literature.
Testable propositions	**T1:** The application of TAG-F reduces TPA turnaround time for the traffic data analyst by reducing or eliminating the need for exploratory data analysis. **T2:** The application of TAG-F leads to higher prediction accuracy for users. **T3:** Innovations adding new PAMs to the corpus of literature lead to prediction accuracy increases for stakeholders applying the framework over time. **T4:** Traffic prediction analytics outcomes added to the corpus of literature lead to prediction accuracy increases for stakeholders applying the framework over time.
Justificatory knowledge (kernel theories)	**J1:** The choice regarding which Predictive Analytics Model (PAM) to adopt in each Traffic Predictive Analytics scenario is fraught with high complexity and is, therefore, prone to uncertainties (Vlahogianni et al. 2014). **J2:** No free lunch (NFL) principle: according to the NFL principle (Wolpert and Macready 1997), there is no single PAM proven to be the best performing in every situation, as demonstrated in Wolpert (2021). A PAM that performs optimally in a given set of conditions or scenarios might perform poorly in other scenarios. Given that traffic prediction takes different forms based on the problem definition, it emphasizes the need to identify which PAM is suitable to be adopted to solve a specific TPA problem. **J3:** Theories underpinning key factors affecting Traffic Prediction (Table 3 in the main body of the manuscript). Theoretical Assumptions: **J4:** Reductionism - Traffic Analytics Guidance design theory can be reduced to simpler parts, and the theoretical underpinnings of these simpler parts will enable an understanding of complex guidance phenomena.

Principles of Implementation	Process Implementation of the TAG-Framework 1. Elicit the analytical problem input attributes from the end user (**PS**) 2. Elicit the domain knowledge input attributes from the end user (**DN**) 3. Elicit the TPA Expertise input attributes from the end user (**TPA_E**) 4. Establish the degree of guidance needed for the end user 5. Execute the TAG-F Guidance Algorithm: Guidance (*Input* (**PS**, **DN**, T**PA_E**), *Analytical Requirements* (**DC**, **DCM**, **PAM**)) → *Output* Guidance Algorithm for Traffic Data Analyst with Intermediate Expertise: 1. Prompt the end user for input on TAG-F dimension values (PS, DN, TPA_E). 2. Access the literature-driven knowledge base to extract meta-knowledge about candidate predictive models (PAMs). 3. Perform a ranking of PAMs based on input TAG-F dimension values. 4. Display a ranked list of recommended PAMs to the end user. 5. Update literature-driven knowledge base (Feedback Loop implemented either via LBD or via Live Project Results)

Expository instantiation	Instantiated artifact is described in Sections 5 and 6 of the main body of the manuscript.

References

Gregor S, Jones D (2004) The formulation of design theories for information systems. Linger H, Fisher J, Wojtkowski W, Wojtkowski WG, Zupančič J, Vigo K, Arnold J, eds., *Constructing the Infrastructure for the Knowledge Economy*, 83–93 (Boston, MA: Springer US), ISBN 978-1-4757-4852-9.

Gregor S, Jones D (2007) The anatomy of a design theory. *Journal of the Association for Information Systems* 8(5)(5).

Henry S, McInnes BT (2017) Literature based discovery: models, methods, and trends. *Journal of biomedical informatics* 74:20–32.

Kasper GM (1996) A theory of decision support system design for user calibration. *Information Systems Research* 7(2):215–232.

Simon H (1981) The sciences of the artificial 2nd edition the mit press. *Cambridge, MA* .

Vlahogianni EI, Karlaftis MG, Golias JC (2014) Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies* 43:3–19.

Walls JG, Widmeyer GR, El Sawy OA (1992) Building an information system design theory for vigilant eis. *Information systems research* 3(1):36–59.

Wolpert DH (2021) What is important about the no free lunch theorems? *Black Box Optimization, Machine Learning, and No-Free Lunch Theorems*, 373–388 (Springer).

Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. *IEEE transactions on evolutionary computation* 1(1):67–82, URL `http://dx.doi.org/10.1109/4235.585893`.