

V Workshop de Dissertações de Mestrado do PPgSI (2018)

FICHA DE PESQUISA

DADOS GERAIS				
Título do projeto de pesquisa	Uso de aprendizado supervisionado para análise de confiabilidade de dados de <i>crowdsourcing</i> sobre posicionamento de ônibus			
Orientando	Diego Vieira Neves			
Orientador(es)	Dr. Daniel de Angelis Cordeiro			
Semestre no curso, na data do workshop	<input type="checkbox"/> 2º semestre	<input type="checkbox"/> 3º semestre	<input type="checkbox"/> 4º semestre	<input checked="" type="checkbox"/> 5º semestre
Qualificação	<input checked="" type="checkbox"/> Qualificação já realizada em: 28/06/2017 <input type="checkbox"/> Realização da qualificação planejada para: dd/mm/aaaa			
Defesa	Prazo máximo para depósito: 28/08/2018 Realização da defesa planejada para: dd/mm/aaaa			
Linha e Área de pesquisa	Gestão e Desenvolvimento de Sistemas: <input type="checkbox"/> BD <input checked="" type="checkbox"/> Engenharia de Software <input type="checkbox"/> Gestão de TI <input type="checkbox"/> IHC		Inteligência de Sistemas: <input type="checkbox"/> IA <input type="checkbox"/> Processamento Gráfico <input type="checkbox"/> Reconhecimento de Padrões	
Área de aplicação	<input type="checkbox"/> Ambientes corporativos <input type="checkbox"/> Bioinformática <input type="checkbox"/> Biometria <input type="checkbox"/> Dispositivos móveis <input type="checkbox"/> Educação	<input type="checkbox"/> Educação a distância <input type="checkbox"/> Governo eletrônico <input type="checkbox"/> Internet <input type="checkbox"/> Jogos <input type="checkbox"/> Jogos sérios	<input type="checkbox"/> Língua Natural <input type="checkbox"/> Linguística <input type="checkbox"/> Processos de Negócio <input type="checkbox"/> Quimioinformática	<input type="checkbox"/> Redes Sociais <input type="checkbox"/> Robótica <input type="checkbox"/> Saúde <input checked="" type="checkbox"/> Smart Cities
Publicações associadas ao projeto de mestrado	Título do artigo: Uso de aprendizado supervisionado para análise de confiabilidade de dados de crowdsourcing sobre posicionamento de ônibus. Autores: Diego Vieira Neves, Felipe Cordeiro Alves Dias e Daniel Cordeiro. Estado: Aceito. Título da conferência: WBCI 2018 – 1º Workshop Brasileiro de Cidades Inteligentes, 2018. Tipo de publicação: Artigo completo			

DESCRIÇÃO DO PROJETO DE PESQUISA	
Contextualização / motivação	Este projeto investiga como o uso de dados obtidos por colaboração coletiva (<i>crowdsourcing</i>) podem aumentar a confiabilidade das informações sobre o transporte coletivo disponibilizados ao usuário do município de São Paulo.
Problema de pesquisa	Esta pesquisa pretende analisar a viabilidade da utilização dos dados obtidos por meio de técnicas de <i>crowdsourcing</i> no enriquecimento dos dados sobre o transporte coletivo fornecidos pelas aplicações de SIU.
Objetivo geral da pesquisa	Esse trabalho propõe um modelo de análise de confiabilidade dos dados obtidos com <i>crowdsourcing</i> que combina métricas escolhidas para o contexto do modal ônibus com técnicas de aprendizagem de máquina para identificar de forma eficiente registros que, por algum motivo, apresentam informações não confiáveis (ausentes ou erradas).
Trabalhos relacionados	O uso de dados de <i>crowdsourcing</i> em problemas de mobilidade urbana já foi investigado pelos autores Pedersen et al. 2013, Misra et al. 2014, Cullina et al. 2015 e Cerotti et al. 2016. Tais estudos se centraram no problema de como construir Sistemas de Informação ao Usuário (SIU) eficientes, confiáveis e em como permitir uma melhor interação entre usuários e gestores de serviços de transporte público. Nenhum dos trabalhos mencionados, entretanto, preocupou-se com a qualidade e confiabilidade dos dados fornecidos pelos usuários.
Proposta para solução	O modelo proposto tem como objetivo realizar, de forma automatizada, a análise de confiabilidade dos dados fornecidos pelos usuários (com <i>crowdsourcing</i>) sobre o horário de passagem dos ônibus nos pontos de paradas. Para cada registro informado por um usuário, o modelo deverá ser capaz de identificar inconformidades e/ou anomalias que possam influenciar ou comprometer a qualidade dos dados coletados.
Dados	Os dados utilizados foram obtidos por uma parceria com a startup Scipopulis, e combinados com os dados que são disponibilizados publicamente pela São Paulo Transporte S. A. (SPTrans).
Validação	A validação será feita por meio da análise dos resultados obtidos, a fim de verificar se as hipóteses foram de fato comprovadas. Outro aspecto desta avaliação será a comparação de soluções propostas por outros autores.
Limitações	Um conjunto de dados com informações insuficiente para realização das atividades de predição.
Resultados esperados	Contribuições científicas: o modelo desenvolvido, poderá auxiliar os pesquisadores da área de análise de dados de mobilidade urbana, dentro do contexto de análise de confiabilidade dos dados de <i>crowdsourcing</i> e predição de horários de chegada de ônibus.

Contribuições tecnológicas: espera-se que o modelo desenvolvido ajude na análise de confiabilidade de dados de mobilidade urbana, contribuindo significativamente no processo de melhoria da qualidade dos serviços disponibilizados para a população.

MÉTODO DE PESQUISA [para cada item, marque todas as opções adequadas para seu projeto]

(basear-se em: <http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>)

Gênero	<input type="checkbox"/> Pesquisa teórica	<input type="checkbox"/> Pesquisa prática	<input type="checkbox"/> Pesquisa empírica	<input type="checkbox"/> Pesquisa metodológica
Natureza	<input type="checkbox"/> Pesquisa básica/pura	<input checked="" type="checkbox"/> Pesquisa aplicada		
Objetivo	<input checked="" type="checkbox"/> Pesquisa exploratória	<input type="checkbox"/> Pesquisa descritiva	<input type="checkbox"/> Pesquisa explicativa	<input type="checkbox"/> [outro – escrever]
Abordagem	<input type="checkbox"/> Pesquisa quantitativa	<input type="checkbox"/> Pesquisa qualitativa	<input checked="" type="checkbox"/> Pesquisa mista (quali-quant)	
Procedimento técnico principal	<input checked="" type="checkbox"/> Pesquisa experimental	<input type="checkbox"/> Pesquisa com <i>survey</i>	<input type="checkbox"/> Pesquisa netnográfica	
	<input checked="" type="checkbox"/> Pesquisa bibliográfica	<input type="checkbox"/> Estudo de caso	<input type="checkbox"/> Teoria fundamentada em dados (<i>Grounded theory</i>)	
	<input type="checkbox"/> Pesquisa documental	<input type="checkbox"/> Pesquisa participante	<input type="checkbox"/> Ciência do projeto (<i>Design science</i>)	
	<input type="checkbox"/> Pesquisa <i>ex-post-facto</i>	<input type="checkbox"/> Pesquisa-ação	<input type="checkbox"/> [outro – escrever]	
	<input type="checkbox"/> Pesquisa de levantamento	<input type="checkbox"/> Pesquisa etnográfica		
Técnica(s) / Instrumento(s) de coleta de dados	<input checked="" type="checkbox"/> Medição		<input checked="" type="checkbox"/> Observação (direta / participante)	
	<input type="checkbox"/> Questionário	<input type="checkbox"/> Formulário	<input type="checkbox"/> Diário de campo / notas de campo	
	<input type="checkbox"/> Entrevista	<input type="checkbox"/> <i>Benchmark</i>	<input type="checkbox"/> Análise documental (ou de artefatos)	
	<input type="checkbox"/> Grupos focais		<input type="checkbox"/> [outro – escrever]	
Técnica(s) de análise de dados	Análise quantitativa:		Análise qualitativa:	
	<input checked="" type="checkbox"/> Estatística descritiva		<input checked="" type="checkbox"/> Análise de conteúdo	
	<input type="checkbox"/> Estatística inferencial		<input type="checkbox"/> Análise do discurso	
	<input type="checkbox"/> [outro – escrever]		<input type="checkbox"/> [outro – escrever]	

CRONOGRAMA [altere conforme necessário]

	2017											2018												
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Estudo da literatura	X	X	X	X																				
Análise da base de dados utilizada		X	X	X	X																			
Levantamento das técnicas mais utilizadas para predição			X	X	X																			
Redação do texto para a qualificação	X	X	X	X	X																			
Exame de Qualificação						X																		
Definição de métricas para identificar dados inconsistentes					X	X	X	X	X	X	X													
Implementações de técnicas						X	X	X	X	X	X	X	X											
Avaliação dos resultados							X	X	X	X	X	X	X											
Redação de artigos científicos												X	X	X										
Redação de texto para a dissertação												X	X	X	X	X	X	X						
Depósito da dissertação																				X				
Defesa da dissertação																						X		

As bibliografias abaixo, ou correlatas, são bastante úteis para lhe ajudar a preencher esta ficha:

- Gerhardt, T. E. & Silveira, D. T. *Métodos de Pesquisa*. EAD – Série Educação a Distância. Editora UFRGS, 2009. Disponível em: <http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>
- Wazlawick, R. S. *Metodologia de Pesquisa para Ciência da Computação*. 2ª. Ed. Elsevier, 2014.
- Creswell, J. W. *Projeto de pesquisa: métodos qualitativo, quantitativo e misto*. 3ª. Ed. Bookman/Artmed, 2010.