Universidade Federal de Santa Catarina

Centro de Araranguá

Programa de Pós-Graduação em Tecnologias da Informação e Comunicação

**DEFESA DE MESTRADO**

**Aluno (a): LUCIANA REGINA BENCKE**

Orientador (a): CRISTIAN CECHINEL

**Data: 01/04/2019** Horário: 14:00 Local: UFSC MATO ALTO Sala: 201

**Título:**

**CLASSIFICAÇÃO AUTOMÁTICA DE MENSAGENS DE REDES SOCIAIS EM DIMENSÕES DOS MODELOS DE CIDADES INTELIGENTES**

**Resumo:**Uma cidade inteligente pode ser definida como uma cidade de alta tecnologia com vários recursos para resolver ou mitigar problemas normalmente gerados pela rápida urbanização. Diferentes modelos de indicadores foram desenvolvidos para acompanhar a evolução das cidades na busca por tornarem-se Cidades Inteligentes. Um exemplo é o padrão 37120 da Organização Internacional para Padronização (ISO), que propõe um conjunto de dimensões e indicadores para serviços e qualidade de vida para cidades e comunidades sustentáveis. Tem sido comum encontrar nas redes sociais perfis oficiais de organizações e entidades governamentais relacionadas aos serviços que elas fornecem ou pelos quais são responsáveis (água, resíduos, transporte, eventos culturais, etc.). Os cidadãos interagem com estes perfis diretamente para comunicar problemas sobre os serviços da cidade. O presente trabalho propõe aplicar algoritmos de aprendizado de máquina sobre os dados urbanos gerados pelas redes sociais, a fim de criar classificadores para categorizar automaticamente as mensagens dos cidadãos de acordo com as diferentes dimensões dos serviços das cidades. Para tanto, dois conjuntos distintos de textos em português foram coletados de duas redes sociais: Twitter (1.950 tweets) e Colab (65.066 postagens). Os textos foram mapeados de acordo com as diferentes categorias ISO 37120, pré-processados e minerados através do uso de 11 algoritmos implementados no Scikit-Learn. Os primeiros resultados indicaram a viabilidade da proposta, com os modelos alcançando médias em torno de 59% para a F1-macro e 75% para a F1-micro ao usar LSVC (Linear Support Vector Classification) e CNB (Complement Naive Bayes). No entanto, como os conjuntos de dados estavam altamente desbalanceados, os desempenhos dos modelos variam significativamente para cada categoria ISO, com os melhores resultados ocorrendo para Transporte (87%), Energia (83%) e Água e Saneamento (73%). Os classificadores gerados neste trabalho podem ser integrados à diversos serviços e sistemas da cidade, tais como: sistemas de suporte à decisão governamental, sistemas de reclamações para cidadãos, painéis comunitários, centrais de polícia, empresas de transporte, produtores culturais, agências ambientais e empresas de reciclagem.

**Palavras-chave**: Classificação de tópicos. Aprendizagem de Máquina. Redes Sociais. Serviços para Cidades Inteligentes. ISO 37120

**Banca examinadora:**

Prof. Cristian Cechinel, Dr.

Orientador

Universidade Federal de Santa Catarina

Prof. Alexandre Leopoldo Gonçalves, Dr.

Membro do PPGTIC

Universidade Federal de Santa Catarina

Prof.ª Merisandra Côrtes de Mattos Garcia, Dr.ª

Membro Externo ao PPGTIC

Universidade do Extremo Sul Catarinense

Prof.ª Silvia Modesto Nassar, Dr.ª

Membro Externo ao PPGTIC

Universidade Federal de Santa Catarina

Prof. Vilson Gruber, Dr.

Membro do PPGTIC (Suplente)

Universidade Federal de Santa Catarina

Prof. Vinicius Faria Culmant Ramos, Dr.

Membro Externo ao PPGTIC (Suplente)

Universidade Federal de Santa Catarina

Obs: É obrigatório o envio deste arquivo (.doc por e-mail: ppgtic@contato.ufsc.br) com antecedência de vinte dias.