



Universidade Federal de Pernambuco
Ciências da Computação
Centro de Informática
2015.1

Proposta de Trabalho de Graduação

**Previsão De Demanda Em Um Sistema De
Compartilhamento De Bicicletas**

Aluno: Rafael Felipe Nascimento de Aguiar
Orientador: Prof. Germano Crispim Vasconcelos

Recife, 12 de Abril de 2015



1. Introdução

Em meados dos anos 2000 surge o termo “Economia Compartilhada”, nomeando um fenômeno que pretendia mudar a maneira como os seres humanos utilizavam recursos [1]. O movimento de Economia Compartilhada questionava a relação quase que linear entre crescimento populacional e a produção de bens duráveis; onde um carro, por exemplo, era produzido com o fim de servir a um único consumidor.

Esse modelo de produção com finalidade individual acabara por prejudicar (e.g., congestionamentos, poluição) meios que são inerentemente compartilhados como as vias de transporte público.

Como exemplo bem sucedido de antagonismo a esse modelo individualista de produção, surgem as primeiras redes de bicicletas compartilhadas; oferecendo uma alternativa mais econômica, ecologicamente correta e benéfica à saúde que o ônibus ou o carro.

Uma rede de bicicletas compartilhadas é uma malha de estações conectadas por rodovias ou ciclovias, em que um usuário pode retirar um bicicleta em um ponto A e devolver em um ponto B [2]. Esses trajetos costumam ser de curta duração e se dividem em duas categorias: lazer e trabalho.

O problema mais desafiador de se manter uma rede como essa é lidar com a distribuição irregular de bicicletas entre estações. Posto que bicicletas podem ser retiradas em uma estação e retornadas em qualquer outra, em um dado instante de tempo estações podem estar tanto vazias quanto cheias; e esses dois estados são problemáticos para rede pois impedem que um usuário inicie uma viagem por falta de bicicletas ou termine uma viagem por falta espaço para retornar uma bicicleta.

Por essa razão, bicicletas precisam ser movidas de estações com alta lotação para estações com baixa lotação e este trabalho aborda uma das formas de se apoiar um sistema de logística para este fim através da previsão de demanda de bicicletas nas estações.



2. Objetivos

O objetivo do presente trabalho é desenvolver um modelo de previsão de demanda de bicicletas em uma rede de compartilhamento considerando dados das estações de compartilhamento (e.g, assinatura de tempo, número de registros) e da meteorologia da região (e.g., temperatura, umidade, velocidade do vento).

A pergunta a qual o modelo deve responder é: dada uma data e hora do dia, qual o número total de bicicletas retiradas?

Esta problemática foi proposta por um popular site de competições na temática de aprendizagem de máquina (Machine Learning - ML) chamado Kaggle¹. Os dados são provenientes da plataforma americana de compartilhamento de bicicletas Capital Bikeshare² e são referentes as retiradas no período de 2011 à 2012 em Washington, DC.

¹<https://www.kaggle.com/c/bike-sharing-demand>

²<http://www.capitalbikeshare.com>



3. Métodos de Implementação

Este trabalho se propõe a implementar tal modelo utilizando técnicas de aprendizagem de máquina.

A justificativa para tal escolha se dá através da análise dos seguintes fatores do problema: não existe uma função que dada um exemplo retorna o resultado exato; existem dados para serem explorados; existem padrões nesses dados.

A aprendizagem de máquina engloba uma série de técnicas que a princípio são divididas em supervisionadas e não supervisionadas. A principal diferença entre elas sendo que a primeira demanda rótulos (labels) nos dados e a segunda não.

Os problemas a serem resolvidos por tais técnicas também comumente se enquadram em duas categorias: classificação e regressão. O primeiro sendo um problema onde dado um item a pergunta a ser respondida é a qual classe esse item pertence (e.g., dado um paciente e seus sintomas, responder se ele é classificado como alto risco ou baixo risco), o segundo sendo um problema onde o objetivo é quantificar o valor de uma variável (dados atributos de uma casa, prever seu valor de mercado).

A previsão de demanda que é objeto deste trabalho se enquadra na categoria de problemas de regressão e o conjunto de dados fornecido possui exemplos com labels para a variável de interesse (a ser medida), possibilitando a utilização de técnicas supervisionadas.

Mesmo depois dessa segmentação do problema ainda existem uma plethora de técnicas que podem ser empregadas; cada técnica explora diferentes conceitos e por consequência tem melhores resultados para certos tipos de dados. Por isso, é comum que um problema seja abordado por diversas técnicas e que a solução final seja obtida através de uma extensiva análise comparativa de resultados.

O presente trabalho se reservará à análise do problema através de um subconjunto reduzido dessas técnicas: redes neurais, random forests e modelos lineares generalizados.



4. Cronograma

2015				
Março	Abril	Maio	Junho	Julho
Especificação do Problema	Avaliação de diferentes modelagens de dados	Otimizar parâmetros	Otimizar parâmetros	Comparar minha técnica com a escolhida como referência para a competição
Análise Exploratória dos Dados	Avaliação de diferentes técnicas de resolução	Analisar o porquê do desempenho de cada técnica e tecer comparativos	Preparar Apresentação	Apresentação
Preparação do Pré-projeto	Entrega do Pré-Projeto	Iniciar Monografia	Ajustes na Monografia	Entrega da Monografia

5. Referências

[1] Benkler, Yochai (2004). "Sharing Nicely: On Shareable goods and the emergence of sharing as a modality of economic production". *The Yale Law Journal* **114**. Retrieved 9 July 2014.

[2] Susan Shaheen and Stacey Guzman (Fall 2011). "Worldwide Bikesharing". *Access Magazine No. 39*. University of California Transportation Center. Retrieved 2012-07-01.



6. Assinaturas

Rafael Felipe Nascimento de Aguiar
Orientando

Prof. Germano Crispim Vasconcelos
Orientador