

DATA SCIENCE: PRÁTICA NO PROJETO AULA ABERTA CESUSC

Luiz Claudio Dalmolin 1
Rejane Costa 3
Sabrina Hanich 5
Lucas Pinheiro 7

Felipe Vieira Domingues Carneiro 2
Sérgio Murilo Schütz 4
André Saibro 6
Guilherme Victorino Martins 8

INTRODUÇÃO

A ciência de dados, chamada de a profissão mais atraente pela Harvard Business Review, é uma área com potencial gigantesco que vem apresentando um aumento de seguidores surpreendente com o passar dos anos, mas que também apresenta poucos especialistas até o momento. De maneira simples, a ciência de dados faz uso da matéria prima (dados) para criar hipóteses que serão comprovadas através de trabalho estatístico e matemático. Podemos dizer que sua finalidade é transformar um dado em informação e a partir dessa informação auxiliar a tomada de decisão para a área na qual está sendo aplicada. Usualmente a ciência de dados está relacionada com três áreas: computação, para a coleta e processamento de dados, estatística, para o estudo da informação adquirida e negócios, para aplicação do conhecimento gerado a fim de impulsionar a empresa em questão.

Por ser considerada uma das tecnologias em ascensão e com maior desatualização, o tema Data Science foi inserido como um dos temas no projeto de extensão "Aula Aberta" do primeiro semestre de 2018. O Projeto Aula Aberta do Curso de Análise e Desenvolvimento de Sistemas (ADS) da Faculdade CESUSC tem como propósito integrar os alunos do Curso de Graduação com profissionais da área de Análise e Desenvolvimento de Sistemas por meio de encontros presenciais no Link Lab na ACATE, promovidos para integrar teoria e prática (SCHÜTZ et al., 2017). A interdisciplinaridade e a Multidisciplinaridade são premissas da essência do curso de ADS, desta forma a construção de atividades que direcionam estes paradigmas trazem enriquecimento e qualidade ao curso, desenvolvendo um senso crítico e inovador aos alunos e professores. Nesta edição 2018/1 a proposta se faz num Hackathon semestral na divisão de 4 equipes das áreas de IOT - Internet das Coisas, Data Science, Mobile e Machine Learning com Devops. Neste pôster será desenvolvido a proposta de trabalho da equipe Data Science.

Figura 1 – Equipe e áreas de Conhecimento



Fonte: autores (2018)

OBJETIVO

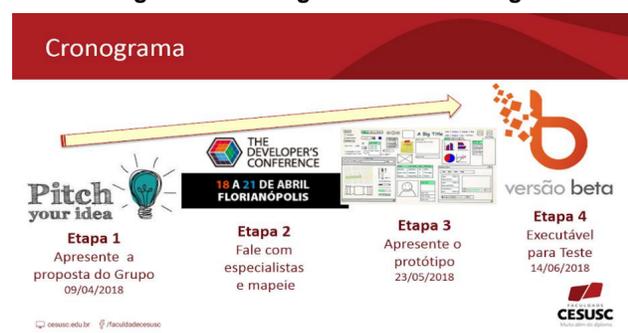
Para tornar possível os passos necessários para a realização de uma análise de dados desde o seu início até o resultado final, optou-se pela seleção de um assunto no qual os dados teriam fácil acesso e que pudessem ser usados sem complicações. O objetivo final é criar uma ferramenta para apresentar a demanda calculada para um par de origem e destino que ainda não possui linha existente, com base em dados socioeconômicos da Fonte ANTT - Viagens Regulares de linhas regulares por meio da Análise de dados transporte rodoviário de passageiros.

METODOLOGIA

Dando continuidade aos temas de aprendizagem desenvolvidos durante o ano de 2017 (Inovação, Geração de Ideias, Modelagem de Negócios, Gestão de Projetos com Canvas, Modelagem de Sistemas), introduzindo áreas de tecnologia e negócios emergentes do mercado (IOT, Data Science, Devops e Machine Learning e Mobile), promovendo por meio de uma competição entre equipes (Hackathon interno do curso de ADS) o desenvolvimento de habilidades de trabalho colaborativo para a construção de produtos/serviços voltados a TI.

Dentro da equipe de Data Science foi realizado as seguintes etapas para a busca do objetivo: Seleção da região de interesse; Filtro de dados para a região de interesse; Levantamento de dados (população, PIB hospitais, turismo); Classificação entre capitais; Processamento dos dados para trazer para o mesmo conjunto de dados; Pesquisas e estudo sobre métodos estatísticos. Para atingir o objetivo deste projeto, será necessário o estudo de técnicas de regressão linear múltipla, que descreve de maneira razoável as relações entre várias variáveis explicativas de um determinado processo. Além disso, faremos a análise de aplicações do mesmo gênero no software Excel da Microsoft, o qual acreditamos que facilitará o relacionamento dos dados arrecadados.

Figura 2 – Cronograma da Metodologia



Fonte: autores (2018)

DESENVOLVIMENTO

Dentro das etapas descritas na metodologia foi realizado estudos direcionados para uma análise da técnica de regressão linear múltipla com exemplos de aplicação no Excel para a identificação da viabilidade de uso para a atividade fim do projeto. Os resultados intermediários esperados serão o levantamento de dados de saúde, levantamento de dados de PIB – POPULAÇÃO e planilhas processadas com a montagem do arquivo com todas as variáveis, sendo aplicado o modelo de regressão linear múltipla produzindo um modelo para calcular uma nova demanda. A proposta produzirá uma ferramenta para apresentar a demanda calculada para um par Origem - Destino que ainda não possui linha existente, apresentando uma análise estatística do mercado, por meio de um portal que gera a demanda calculada. A regressão linear múltipla é uma técnica multivariada cuja finalidade principal é obter uma relação matemática entre uma das variáveis estudadas (variável dependente ou resposta) e o restante das variáveis que descrevem o sistema (variáveis independentes ou explicativas), e reduzir um grande número de variáveis para poucas dimensões com o mínimo de perda de informação, permitindo a detecção dos principais padrões de similaridade, associação e correlação entre as variáveis. (PESTANA & GAGEIRO, 2005)

CONSIDERAÇÕES FINAIS

Este projeto de aula aberta está permitindo inserir no meio acadêmico áreas de Tecnologia em expansão no mercado de trabalho, como neste caso Data Science, desenvolvendo nos alunos as habilidade em Inovação, geração de Ideias, prototipagem e desenvolvimento de código, gerenciamento de projetos na prática, integração e comunicação com profissionais do mercado. O estudo dos dados estão sendo analisados com base nas técnicas e ferramentas descritas da área Data Science na busca de melhorias para a mobilidade urbana.

1 Mestre e Professor do CST em ADS. Faculdade Cesusc / e-mail: dalmolin@labtrans.ufsc.br
2 Graduando em CST em ADS. Faculdade Cesusc/ e-mail: felipevdcarneiro@gmail.com
3 Mestre e Professora do CST em ADS. Faculdade Cesusc / e-mail: rejanecestafloripa@gmail.com
4 Mestre Professor e Coordenador do CST em ADS. Faculdade Cesusc / e-mail: sergio.schutz@cesusc.edu.br
5 Graduando em CST em ADS. Faculdade Cesusc / e-mail: hanich.sabrina@gmail.com
6 Graduando em CST em ADS. Faculdade Cesusc / e-mail: andre_h_saibro@hotmail.com
7 Graduando em CST em ADS. Faculdade Cesusc / e-mail: lucascpinheiro.studies@gmail.com
8 Graduando em CST em ADS. Faculdade Cesusc / e-mail: guivmartins@gmail.com

REFERÊNCIAS BIBLIOGRÁFICAS

PESTANA, Maria Helena; GAGEIRO, João Nunes. **Análise de Dados para ciências sociais : complementariedade do SPSS**. 6. ed. Lisboa: Sílabo, 2014.
SCHUTZ, Sérgio Murilo; Dias, Ibsem; LENGLE, Fernando; FERNADES, Roberto Fabiano. **Projeto de Extensão Sala de Aula Aberta**. 1. ed. Florianópolis: CESUSC, 2017