Aprendizagem de máquina: Regressão Linear em uma instituição de Ensino

Caio Vinicius Batista Martins, Lyncon Rodrigo de Melo Santos, Felipe Vieira Mota

Faculdade de Tecnologia de São Vicente - FATEF

Email: cvmartins1996@gmail.com

Resumo: O presente artigo apresentará um estudo de caso em uma instituição de ensino empregando técnicas de Aprendizado de Máquina a fim de demonstrar a relação existente entre um curso superior e a inserção dos jovens no mercado de trabalho no universo da Faculdade de Tecnologia de São Vicente exclusivamente para os cursos de Pedagogia e Administração.
Palavras chave: Aprendizado de máquina, inserção dos jovens no mercado de trabalho como um graduando de pedagogia e administração.

Machine Learning: Linear Regression applied at a Higher-Education Institution

Abstract: This article aims to go through a study case applied at Faculdade de Tecnologia de São Vicente. It hereby envisions to establish the relationship between higher education and getting into the market through Machine Learning algorithms exclusively applied for the graduate students of Pedagogy and Business Management in the study-case institution.
Keywords: Machine Learning, getting into the market as a pedagogy and business management graduate

Introdução

O campo de Ciência de Dados tem crescido como fator crucial para predizer tendências ocultas em dados[5]. Para tanto, o campo utiliza de técnicas estatísticas, dentre elas, Modelos Lineares Aplicados que serão os utilizados neste artigo para demonstrar os modelos encontrados.

Regressão é uma metodologia estatística que utiliza da relação entre duas ou mais variáveis quantitativas a fim de que a variável de resposta possa ser prevista a partir das anteriores[1]. Sua relevância é muito grande (conforme uma vez mais publicado pela Rede Linux IME-USP por Carlos ALberto Ynoguti).[4]

A fim de expor a aplicação de modelos lineares, aplicamos uma pesquisa de campo com os alunos da Faculdade de Tecnologia de São Vicente - FATEF. O total em porcentagem de alunos é dado conforme Figura 1. Do conjunto dos cursos oferecidos na instituição, foram priorizados a amostragem de alunos dos cursos de Administração e Pedagogia visto a facilidade

de entrada no mercado de trabalho conforme Fundação do Desenvolvimento Administrativo do Estado de São Paulo.[3]

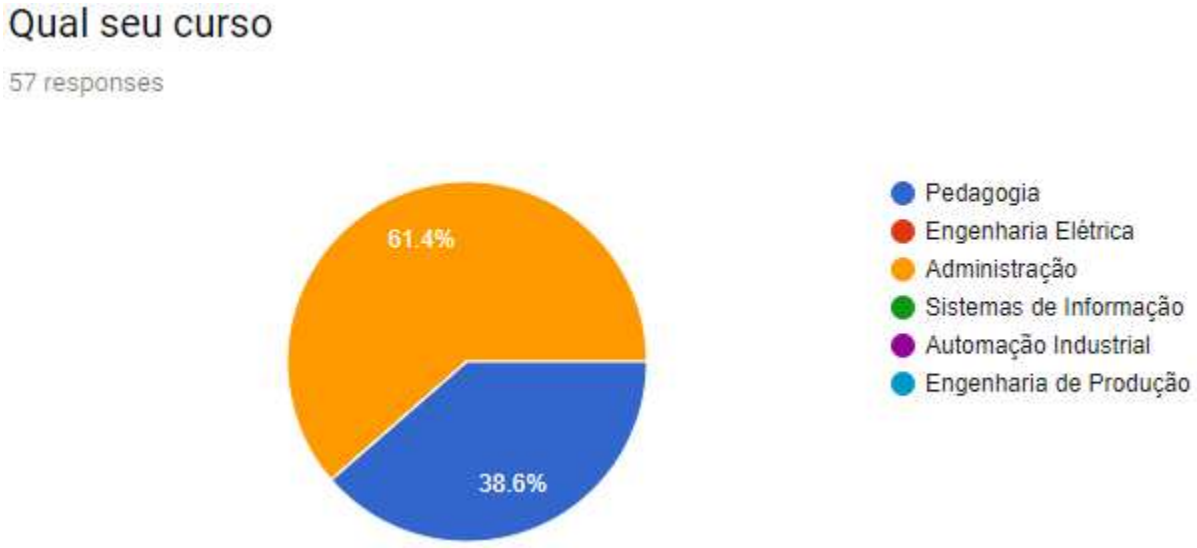


Figura 1. Porcentagem de alunos entrevistados

Após levantados os dados sobre a relação de pessoas que trabalham na mesma área de atuação do curso (conforme Figura 2) foi identificado um padrão através da alimentação do algoritmo de Regressão Linear produzido em Python.



Figura 2. Porcentagem de alunos entrevistados que trabalham ou não na área

Objetivos

O presente trabalho tem como objetivo instaurar a predição dos dados computados por meio de aprendizado de máquina em algoritmos aplicados na linguagem Python a partir de

predicados coletados em uma pesquisa de campo na Faculdade de Tecnologia de São Vicente - FATEF.

Material e métodos

A abordagem consistiu numa pesquisa de campo no dia 22 de Setembro de 2017 das 21:00 horas até às 22:15 na faculdade de Tecnologia de São Vicente - FATEF utilizando a ferramenta *Google Forms*, onde foi elaborado um formulário contendo as seguintes questões:

1 - Qual o curso?

2 - Qual o semestre?

3 - Trabalha na área que cursa atualmente?

4 - Se sim, há quantos meses?

A partir das respostas obtidas em formato de tabela CSV, pode-se instaurar o modelo para o aprendizado de máquina: Regressão Linear Simples codificado na linguagem Python.

Como consequência de ser uma regressão linear simples, foram descartados do material outras variáveis, tais como, "se já é formado na área; Se já possui tecnico na area; ou correlacionados", ou como eventos de crise financeira e índices de desemprego.

A partir daqueles que responderam "sim" (que trabalham na área), foi elaborado o seguinte modelo em regressão linear: Sendo abscissa de X semestre em que ela se encontra (convertido de semestre para meses) e ordenadas de Y para mensurar quantos meses possuem de trabalho na área.

Sendo assim, dada a fórmula para regressão linear $Y = B_0 + B_1 * X_1$, obtém-se assim a previsão de um universo matemático esperado, onde X é a variável independente , Y a variável dependente, B_1 representa o declínio da curva, ou seja, representa o coeficiente angular de inclinação da reta e B_0 representa uma constante na equação, onde se $X_1 = 0$ então Y = B_0.

Resultados

Dos que disseram sim (correspondentes aos 63,8% que atuam na área) a média de meses trabalhados foi de 18. A partir disso, foi iniciado o modelo de Machine Learning com regressão linear.

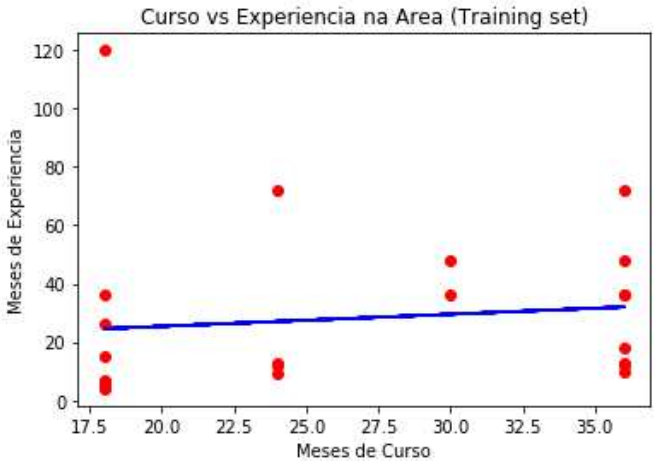


Figura 3. Modelo de treinamento

Observando por exemplo a partir do ponto x = 27,5 correspondente a 24 meses de trabalho, os pontos iniciam uma aproximação gerando um padrão linear. Percebe-se que o modelo de aprendizagem de máquina se comportou muito bem perante as adversidades encontradas tais como ruídos, visto haver encontrado uma linearidade entre os pontos.

Vale ressaltar que foi válida na nossa coleta a contagem de horas para estudantes de pedagogia a modalidade de estágio visto que segundo a LEI Nº 9.394, DE 20 DE DEZEMBRO DE 1996. TÍTULO VI Dos Profissionais da Educação Art. 64 o profissional deve atuar somente após a formação completa.[6]

Conclusões

De acordo com a amostragem coletada e trabalhada visualizamos que mesmo embora haja pontos dispersos observou-se que por volta do 5º semestre houve um aumento na margem de alunos inseridos no mercado de trabalho nos setores de Administração e Pedagogia.

Agradecimentos

Os autores deste artigo gostariam de agradecer o apoio dado pela Faculdade de Tecnologia de São Vicente - FATEF, ao professor Mestre João Carlos pela indicação e especialmente ao professor Mestre Gilmar Aquino por dispor de tempo e esforço no auxílio do desenvolvimento deste desafio proposto.

Referências Bibliográficas

1. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models.* New York: McGraw-Hill/ Irwin.
2. MELO, S. L. de. BORGES, L. O.. (2007) A transição da universidade ao mercado de trabalho na ótica do jovem. Site: http://www.anpad.org.br/admin/pdf/enanpad2005-gprb-1354.pdf. Acesso em 30/09/2017
3. DOSP - Diário Oficial de São Paulo. (02 de Fevereiro de 2016). *Desenvolvimento Administrativo do Estado de São Paulo.* disponível em Diário Oficial de São Paulo. Site: http://www.saopaulo.sp.gov.br/spnoticias/ultimas-noticias/fundap-abre-inscricoes-para-programa-de-estagio/. Acesso em 30/09/2017
4. Ynoguti, C. A. (2011). *Probabilidade, Estatística e Processos Estocásticos.* A Rede Linux do IME-USP. Site: https://www.linux.ime.usp.br/~daniloss/antes-2012/Apostila%20TP501%20-%20Ynoguti,%202011.pdf. Acesso em 30/09/2017.
5. VASANT DHAR (2013). Data Science and Prediction. Ca' Foscari Universidade de Veneza na Itália. Site: http://venus.unive.it/romanaz/LM_Data_Science/dhar_datascience_prediction.pdf. Acesso em 30/09/2017
6. LEI Nº 9.394, DE 20 DE DEZEMBRO DE 1996. TÍTULO VI - Dos Profissionais da Educação. Presidência da República Casa Civil Subchefia para Assuntos Jurídicos. Site: http://www.planalto.gov.br/ccivil_03/leis/L9394.htm. Acesso em 30/09/2017.