Projeto de Formatura – Turmas 2018

PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia de Computação

Tema:

TRADING DE AÇÕES COM BASE EM ANÁLISE DE MÍDIAS SOCIAIS

INTRODUÇÃO E MOTIVAÇÃO

O mercado financeiro, principalmente o acionário, é altamente influenciado pela divulgação de notícias. Há casos em que a divulgação de uma notícia relacionada a determinada companhia pode gerar um grande impacto no preço da ação desta companhia, negociada na bolsa de valores.
Muitos modelos que buscam prever o preço de ativos financeiros usam somente dados de mercado (preço, volume negociado, etc.) e outros fazem somente a analise de sentimento de notícias. O uso de Inteligência Artificial junto a dados de mercado e notícias ainda é pouco utilizado no Brasil e é um mercado que apresenta enormes oportunidades.

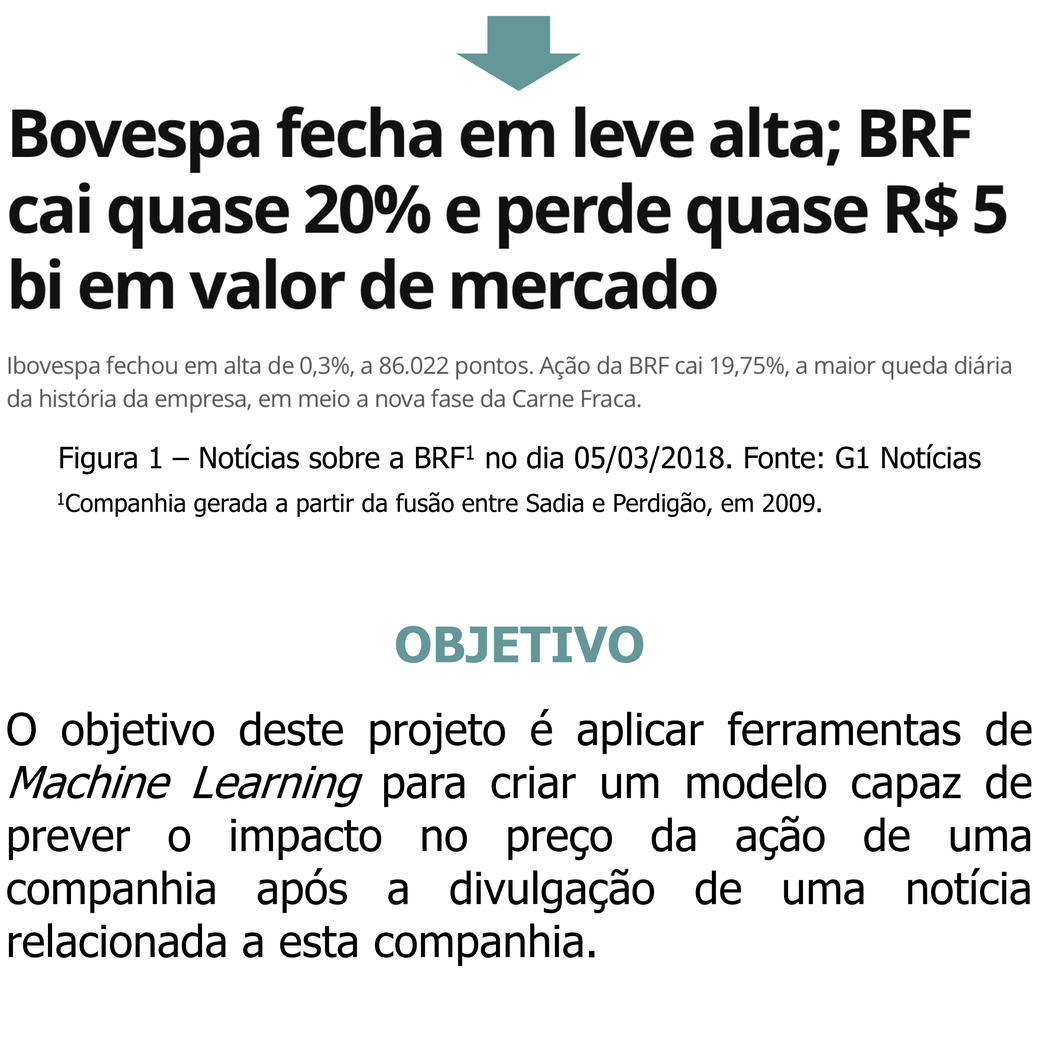

Figura 1 – Notícias sobre a BRF[1] no dia 05/03/2018. Fonte: G1 Notícias

[1]Companhia gerada a partir da fusão entre Sadia e Perdigão, em 2009.

OBJETIVO

O objetivo deste projeto é aplicar ferramentas de *Machine Learning* para criar um modelo capaz de prever o impacto no preço da ação de uma companhia após a divulgação de uma notícia relacionada a esta companhia.

FERRAMENTAS

Integrantes: Henrique Hashimoto Cavalcante
Matheus Henrique de Campos
Thábata Sakurai Goromar

Professor Orientador: Prof. Dr. Jorge Luis Risco Becerra

METODOLOGIA

Para realizar o projeto, utilizamos a metodologia CRISP-DM para entender os dados e depois prepará-los, para só então realizar a modelagem e avaliar a acurácia do método proposto. Depois do teste e validação do modelo, o processo se repete com alteração de parâmetros com o intuito de melhorar os resultados até que se alcancem os valores esperados.

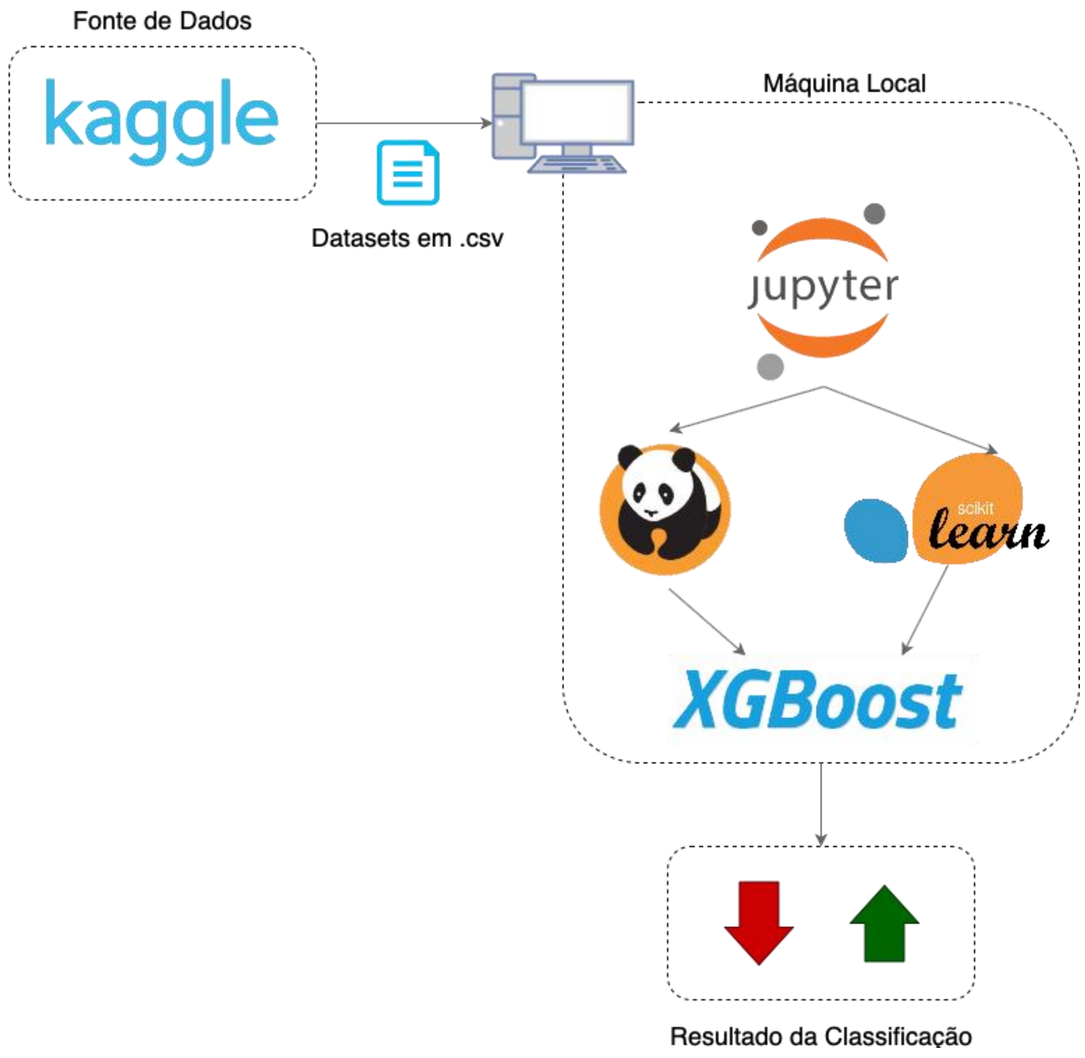

Figura 2 – Visão geral do funcionamento do projeto

RESULTADOS

Foram realizados testes localmente e dentro da plataforma *Kaggle*, realizando modificações nas *features* de entrada, no tratamento de dados e nas parametrizações dos modelos. Os resultados adquiridos se encontram na seguinte tabela:

Modelo	Acurácia	Ambiente
Logistic Regression	54,67%	Kaggle
K Nearest Neighbors	75,89%	Kaggle
Gaussian Naive Bayes	54,15%	Kaggle
Stochastic Gradient Descent	51,76%	Kaggle
Linear Suppot Vector Machine	54,57%	Kaggle
Decision Tree	54,76%	Kaggle
Random Forest	54,93%	Kaggle
XGBoost	54,66%	Kaggle
LightGBM	70,13%	Kaggle
LightGBM	76,81%	Local
XGBoost	80,70%	Local