Faculdade de Ciência e Tecnologia de Montes Claros

Felipe Israel Corrêa

SISTEMA AGREGADOR PARA ANÁLISE DE CARACTERÍSTICAS DE IMÓVEIS COM RASTREAMENTO WEB E APRENDIZADO DE MÁQUINA

Montes Claros- MG
2018

Felipe Israel Corrêa

SISTEMA AGREGADOR PARA ANÁLISE DE CARACTERÍSTICAS DE IMÓVEIS COM RASTREAMENTO WEB E APRENDIZADO DE MÁQUINA

Monografia apresentada ao Curso de Engenharia da Computação, da Faculdade de Ciência e Tecnologia de Montes Claros, como parte dos requisitos para obtenção do título de Engenheiro da Computação.

Orientador: **PROF. DR. RENATO DOURADO MAIA.**

Montes Claros - MG
2018

FUNDAÇÃO EDUCACIONAL MONTES CLAROS
Faculdade de Ciência e Tecnologia de Montes Claros

Felipe Israel Corrêa

SISTEMA AGREGADOR PARA ANÁLISE DE CARACTERÍSTICAS DE IMÓVEIS COM RASTREAMENTO WEB E APRENDIZADO DE MÁQUINA

Esta monografia foi julgada adequada como parte dos requisitos para a obtenção do diploma de Engenheiro da Computação aprovada pela banca examinadora da Faculdade de Ciência e Tecnologia de Montes Claros.

_____
Prof. Dr. Maurílio José Inácio
Coord. do Curso de Engenharia da Computação

Banca Examinadora

Prof. Dr. Renato Dourado Maia, FACIT / (Orientador) _____

Prof. Examinador 1 _____

Prof. Examinador 2 _____

Montes Claros, xx de junho de 2018.

Para Lavínia, amor maior.

AGRADECIMENTOS

Agradeço primeiramente a Deus pela saúde.

Aos meus pais Aroldo e Isolina (*In memorian*) pelos contínuos esforços em me oferecerem uma educação satisfatória.

A minha esposa Laís pelo companheirismo e apoio em toda esta jornada.

Meus irmãos e sobrinhos pela amizade sempre presente.

Aos meus colegas e professores, em especial ao professor Renato Dourado, que como eu, acreditou neste projeto.

“Não se espante com a altura do voo. Quanto mais alto, mais longe do perigo. Quanto mais você se eleva, mais tempo há de reconhecer uma pane. É quando se está próximo do solo que se deve desconfiar. ”

Santos Dumont

RESUMO

A democratização do acesso a internet tem proporcionado uma geração de dados, sobretudo digitais, nunca antes vista na história da humanidade. Nunca se produziu tanta informação em tão pouco tempo. A partir desta perspectiva este estudo tem por objetivo coletar os dados sobre imóveis residenciais, da cidade de Montes Claros, adquiridos em sites de empresas imobiliárias e desenvolver um sistema que os agregue e gere observações relevantes sobre as características destes empreendimentos. A aquisição das informações será feita por meio de um Rastreador Web e a análise dos dados ocorrerá através da aplicação de Aprendizagem de Máquina no que tange a utilização de um algoritmo regressor e um de recomendação e busca. Para alcance do objetivo foram estudados os conceitos que englobam os processos de rastreamento e aprendizagem de máquina, verificadas as condições para desenvolvimento do trabalho e descritos os algoritmos de codificação. Como proposto, o sistema obteve êxito em predizer valores de um imóvel com base em suas características, tais como bairro, número de quartos, banheiros, vagas de estacionamento e tamanho da área, além de efetuar a recomendação e busca de imóveis. Todo ele foi confeccionado através da linguagem de programação *Python* no ambiente de desenvolvimento *Visual Studio Code*, em conjunto com as bibliotecas *Requests* e *BeautifulSoup* para o rastreador e as bibliotecas científicas e gráficas *Scikit-learn*, Pandas, *Numpy*, *Matplotlib* e *Bokeh* para aprendizagem de máquina.

Palavras-Chave: *Dados, Imóveis, Rastreador Web, Aprendizagem de Máquina, Python.*

ABSTRACT

The popularization of Internet access has provided a generation of data, especially digital data, never before seen in the history of mankind. Never has so much information been produced in such a short time. From this perspective, this study aims to collect data on real estate acquired on websites of real estate companies and develop a system that aggregates and generates relevant observations about the characteristics of these developments. The information acquisition will be done through a Web Tracker and data analysis will take place through the Machine Learning application. In order to reach the objective, the concepts that comprise the processes of machine tracking and learning have been studied, the conditions for the development of the work were verified and the coding algorithms were described. As proposed the system is able to predict values of a property based on its characteristics, such as neighborhood, number of rooms, bathrooms, parking spaces and size of the area, in addition to recommending and searching for real estate. The system was made using the Python programming language in the Visual Studio Code development environment, together with the Requests and BeautifulSoup libraries for the tracker, and the Scikit-learn, Pandas, Numpy, Matplotlib and Bokeh scientific and graphic libraries for learning machine. At the end, this work was successful in developing the system.

Keywords: *Data, Real Estate, Web Tracker, Machine Learning, Python.*

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO