

APLICAÇÃO DA TÉCNICA SUPPORT VECTOR MACHINE (SVM) EM UM PROBLEMA DE RISCO DE CRÉDITO¹

Matheus Andrade Angel Lopez², Analucia Vieira Fantin³

¹ Vinculado ao projeto “Machine Learning: estudo de aplicações no setor financeiro”

² Acadêmico do Curso de Ciências Econômicas – ESAG – Bolsista PIBIT/CNPq

³ Orientadora, Departamento de Ciências Econômicas – ESAG – analucia.fantin@edu.udesc.br

Este trabalho teve como objetivo principal utilizar ferramentas de aprendizado de máquina em um problema de risco de crédito.

Inicialmente, o aluno realizou uma extensa pesquisa e estudo autodidata sobre os conceitos fundamentais de Python, incluindo variáveis, estruturas de controle, funções e manipulação de dados. Através de exercícios sugeridos pela orientadora, o aluno aprofundou seu conhecimento nas principais bibliotecas: *Numpy*, *Pandas* e *ScikitLearn*. A plataforma DataCamp (<https://www.datacamp.com/>) ajudou o aluno nesta fase do aprendizado. A plataforma oferece cursos interativos e práticos sobre uma variedade de tópicos relacionados a dados e ciência de dados, incluindo a linguagem de programação Python.

O primeiro curso realizado dentro da plataforma foi o “Introduction to Python”, com o foco no aprendizado da sintaxe da linguagem de programação Python. O segundo curso, “Intermediate Python”, possibilitou ao aluno aprofundar um pouco mais os conceitos e a lógica de programação. O terceiro curso realizado foi o “Data Manipulation with Pandas”, o qual ensina e exercita o uso da biblioteca Pandas, uma das principais bibliotecas em Python para manipulação de dataframes. O último curso do roteiro de aprendizado foi o “Introduction to Statistics in Python”, o qual introduziu conceitos de estatística.

A segunda fase deste projeto envolveu o aprendizado e a aplicação da biblioteca *Scikit-Learn*. *Scikit-Learn* é uma biblioteca Python que fornece uma ampla gama de ferramentas para tarefas de aprendizado de máquina, como classificação, regressão, agrupamento e pré-processamento de dados.

A terceira fase do projeto envolveu o desenvolvimento dos primeiros algoritmos, utilizando as ferramentas disponíveis do Python. Nesta fase o aluno foi desafiado a criar um algoritmo para manipular uma planilha de dados, ou seja, um dataframe. O objetivo do exercício era ler e atualizar os dados da planilha através das seguintes ações: ler os dados da planilha, verificar a existência de um dado, incluir dados novos, excluir dados duplicados e salvar a planilha atualizada em arquivo.

A última fase do projeto ainda previa a aplicação das ferramentas Python e *Scikit Learn* em um problema prático de aprendizado de máquina, desde a pré-processamento de dados até a criação de um modelo de Machine Learning, avaliação de desempenho e interpretação de resultados. Porém, o estudante não teve tempo hábil para iniciar a última fase, visto que o mesmo iniciou suas atividades em Maio/2023, devido a uma troca de bolsista durante o último ano de projeto. Desta forma, o estudante participou ativamente do projeto por apenas 4 meses, finalizando apenas a fase de capacitação.

Palavras-chave: *Machine Learning, Numpy, Pandas, Scikit Learn*