

Demanda da academia da UC Berkeley durante os anos de 2015 e 2017

1st Aline Fortaleza Ferreira da Silva

Centro de Informática

Universidade Federal de Pernambuco

Recife, Brasil

affs2@cin.ufpe.br

2nd Arthur Alves Marsaro

Centro de Informática

Universidade Federal de Pernambuco

Recife, Brasil

aam4@cin.ufpe.br

3rd Davi Gonzaga Guerreiro Barboza

Centro de Informática

Universidade Federal de Pernambuco

Recife, Brasil

dggb@cin.ufpe.br

4th Lucas da Silveira Absalão

Centro de Informática

Universidade Federal de Pernambuco

Recife, Brasil

lsa4@cin.ufpe.br

5th Pedro Victor Saraiva Campello

Centro de Informática

Universidade Federal de Pernambuco

Recife, Brasil

pvs4@cin.ufpe.br

Abstract—This research introduces a methodology for forecasting the optimal time to visit the gym at UC Berkeley by employing statistical analysis of available data. Utilizing a dataset sourced from the Kaggle website, the study concentrates on constructing a predictive model using Python, incorporating the principles of the Naive Bayes classifier, along with other machine learning techniques. The primary objective is to furnish students with a tool that enables them to determine the most advantageous time to access the University gym, taking into account a range of influencing factors.

Index Terms—Demanda, Academia, Naive Bayes, Modelo Preditor

I. INTRODUÇÃO

A prática de atividades físicas e o cultivo de hábitos saudáveis tornaram-se uma preocupação crescente na sociedade contemporânea. Com o aumento do interesse na busca por qualidade de vida, as academias de ginástica se tornaram espaços fundamentais para o desenvolvimento de atividades físicas regulares. No ambiente universitário, esses espaços desempenham um papel crucial na promoção do bem-estar entre os estudantes, professores e demais membros da comunidade acadêmica.

Assim, com uma vida universitária cada vez mais demandante, é primordial para os estudantes conseguirem prever o quanto lotada estará a academia do campus em cada momento do dia. Analisar os fatores do dia pode contribuir para um melhor uso do tempo dos estudantes no dia-a-dia, algo precioso no mundo acadêmico.

II. OBJETIVOS

Temos como finalidade desenvolver um classificador ingênuo de Bayes para estimar a demanda da academia com base em diferentes atributos, como exemplo horário, temperatura, período do mês e etc.

III. JUSTIFICATIVA

Pela demanda atual do público e pela tendência cada vez maior de uma vida mais saudável, nossa proposta tem como objetivo otimizar e analisar como os fatores externos influenciam na quantidade dos usuários de academia durante o dia, o que pode demonstrar que, caso algum evento aconteça, pode haver um crescimento ou diminuição da quantidade de pessoas em um determinado horário. Logo, pretendemos, por intermédio de aplicações matemáticas, permitir encontrar horários ditos como “ideais” para os frequentadores da academia, transformando um local de mais conforto e com um melhor rendimento para os usuários.

IV. METODOLOGIA

Neste projeto, será analisado os diferentes níveis de demanda sobre uma academia fazendo o uso de machine learning. Nesse sentido, utilizaremos o classificador de Bayes ingênuo para a categorização dos graus de demanda e lotação utilizando diferentes variáveis. Com isso usaremos o ambiente virtual do Google Colaboratory com códigos implementados na linguagem computacional python e algumas das bibliotecas disponíveis nela, como Scipy, Numpy, Pandas, Matplotlib, Seaborn e Scikitlearn.

A. Dataset

A base de dados utilizada para fundamentar o projeto encontra-se disponível na plataforma Kaggle, apresentando 9 atributos e 62,184 pessoas que frequentaram a University of California, Berkeley.

Os atributos foram organizados da seguinte maneira:

- 1) Data - String, indicando o dia analisado.
- 2) Hora do Dia - Inteiro [0 - 86,400], indica o período do dia, em segundos, mais frequentado.
- 3) Dia da Semana - Inteiro [0 (segunda) - 6 (domingo)].
- 4) É Final de Semana - Booleano, 0 (dia útil) e 1 (fim de semana).

- 5) É Feriado - Booleano, 1 (feriado) e 0 caso contrário.
- 6) Temperatura - Float, graus fahrenheit.
- 7) É Começo do Período - 1 (início do período universitário) e 0 caso contrário.
- 8) Mês - Inteiro [1 (janeiro) - 12 (dezembro)].
- 9) Hora - Inteiro [0-24].

B. Processamento do dataset

O processamento do dataset é utilizado para transformar as informações presentes na base de dados em um formato útil e eficiente antes da análise. Inicialmente, é feito o processo de criação de atributos, que divide os dados em classes semelhantes, os atributos. Contudo, esse procedimento não se faz necessário no presente projeto, pois as informações já são fornecidas na base de dados do kaggle com suas devidas classificações.

Outra etapa do processamento se trata do filtro de instâncias. Ele é necessário para alterar ou remover as instâncias que apresentam inconsistência, incompletude ou que são dispensáveis à análise. Assim como na criação de atributos, esse processo já foi feito pela própria base de dados.

Posteriormente, se faz a seleção de atributos, a qual parte do princípio que o banco de dados apresenta atributos redundantes e irrelevantes para o trabalho. Dessa forma, são selecionados apenas os dados necessários e que facilitarão o modelo de aprendizagem.

Todas essas etapas se mostram primordiais no processo de engenharia de atributos para que os dados estejam no formato adequado à entrada do algoritmo de aprendizagem. Portanto, será possível fazer as análises necessárias de maneira mais eficiente e apropriada.

C. Teorema de Bayes

Na área de teoria das probabilidades e estatística, o teorema de Bayes, também conhecido como lei de Bayes ou regra de Bayes, aborda a probabilidade de um evento com base em um conhecimento prévio que pode estar associado a esse evento. Esse teorema ilustra como ajustar as probabilidades prévias à luz de novas evidências, resultando em probabilidades posteriores.

O nome "teorema de Bayes" é atribuído ao pastor e matemático inglês Thomas Bayes (1701-1761). Ele dedicou seus estudos ao cálculo da distribuição para o parâmetro de probabilidade em uma distribuição binomial, utilizando terminologia moderna.

Uma aplicação frequente do Teorema de Bayes ocorre em classificadores probabilísticos, como o mencionado Classificador Naive Bayes. Nesse cenário, o teorema de Bayes é empregado para calcular a probabilidade de uma instância pertencer a uma classe específica com base nas características (atributos) observadas.

D. Classificador de Naive Bayes

Um dos algoritmos mais tradicionais e significativos em aprendizado de máquina, o Naive Bayes destaca-se tanto na comunidade acadêmica quanto no mercado. Ele representa

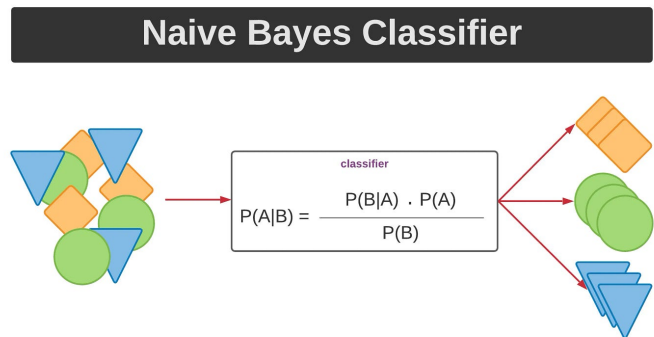


Fig. 1. Classificador de Naive Bayes.

uma abordagem simples para problemas de classificação, proporcionando uma sólida fundamentação estatística para as atividades de aprendizado de máquina (ML).

O classificador Naive Bayes é fundamentado nas descobertas de Thomas Bayes, sendo utilizado para realizar previsões em contextos de aprendizado de máquina. O termo "naive" (ingênuo) refere-se à maneira como o algoritmo analisa as características de um conjunto de dados, assumindo que essas características são independentes entre si.

Adicionalmente, o algoritmo pressupõe que todas as variáveis features são igualmente relevantes para o resultado. Em situações em que essa premissa não se aplica, essa técnica pode não ser a escolha ideal.

Para além de suas contribuições para a estatística, Bayes também desempenhou um papel crucial no universo do aprendizado de máquina. O classificador Naive Bayes figura como um dos principais algoritmos na área, servindo como o primeiro contato com esse domínio para muitas pessoas. A fórmula do classificador define claramente os passos a serem seguidos no processo de aprendizado.

Devido à sua simplicidade e à base estatística acessível, o Naive Bayes é frequentemente preferido em muitos cenários.

E. Aplicações

Nos códigos implementado em python no Google Colab cada biblioteca tem sua função para ajudar no desenvolvimento do projeto. A biblioteca Scikit-learn será utilizada para facilitar o uso do algoritmo Naive Bayes, enquanto bibliotecas como Pandas e Numpy serão responsáveis pela análise de dados e as operações matemáticas nelas utilizadas. Desta forma, a combinação dessas ferramentas viabiliza a exploração completa do Classificador Naive Bayes.

Os dados dentro do projeto serão analisados e tratados para garantir sua qualidade e relevância. Desse modo, o conjunto de dados será dividido, tendo o conjunto de treinamento e o conjunto de teste. O conjunto de treinamento é responsável por alimentar e ensinar o modelo de classificação, assim ajustando os padrões ao que estarão presentes nos dados. Enquanto o de teste conclui a etapa de verificação da eficiência do algoritmo classificador Naive Bayes, garantindo a capacidade de gerar

dados inusitados, assim também de avaliar a suas métricas de desempenho e de realizar previsões precisas.

Desse modo, realizando o objetivo do projeto de prever os cenários para a escolha do momento de se ir na academia, tendo assim o controle probabilístico sobre as piores e melhores ocasiões para se visitar o estabelecimento e adquirir conhecimento das características que mais influenciam na análise.

V. CRONOGRAMA DE ATIVIDADES

O grupo seguirá o plano de atividades proposto pela tabela a seguir:

TABLE I
CRONOGRAMA DE ATIVIDADES

Datas (planejadas)	Atividades	
	Ação	Detalhes
14/12/2023	Seleção do Data Set	Em grupo
16/01/2024	Escrita da Proposta do projeto	Em grupo
29/01/2024	Entrega da Proposta do projeto	Classroom
29/01 - 28/02/24	Desenvolvimento do Projeto	Colab/Python
03/03/2024	Elaboração dos slides	Em grupo
08/03/2024	Finalização do Relatório	Em grupo
11/03/2024	Entrega final do projeto	-

VI. RESULTADOS E DISCUSSÃO

REFERENCES

- [1] Crowdedness at the Campus Gym
<https://www.kaggle.com/datasets/nsrose7224/crowdedness-at-the-campus-gym/data>
- [2] Bussab, Wilton de O.; Morettin, Pedro A. (2010). Estatística Básica 6 ed. São Paulo: Saraiva
- [3] Gelman, Andrew; Carlin, John B.; Stern, Hal S.; Dunson, David B.; Vehtari, Aki; Rubin, Donald B. (2014). Bayesian data analysis 3 ed. Boca Raton: Chapman Hall
- [4] Naive Bayes: Como funciona esse algoritmo de classificação
<https://blog.somostera.com/data-science/naive-bayes>
- [5] Classificador Naive Bayes na biblioteca Scikit Learn
<https://scikit-learn.org/stable/modules/naivebayes.html>