

Reconhecimento de Dígitos

Arthur Abrahão Santos Barbosa
Universidade Federal de Pernambuco
Centro de Informática
Pernambuco, Brasil
aasb2@cin.ufpe.br

Filipe Samuel da Silva
Universidade Federal de Pernambuco
Centro de Informática
Pernambuco, Brasil
fss8@cin.ufpe.br

I. Objetivos

A. Objetivo Geral

Através da análise de algumas informações referentes a um indivíduo, usando um classificador ingênuo de Bayes, prever se o mesmo irá se inscrever em um depósito a prazo.

B. Objetivos Específicos

- Compreender a implementação do classificador ingênuo de Bayes
- Demonstrar a Importância do Aprendizado de máquina e suas aplicações
- Investigar o uso do Aprendizado de máquina no marketing bancário

II. Justificativa

Este projeto foi escolhido com base na maneira organizada e completa que o conjunto de dados foi disponibilizado e por sua afinidade em aplicar-se os conceitos existentes, o banco de dados pode ser encontrando do site Machine Learning Repository, com o nome de "Bank Marketing Data Set" [1].

Sua função é prover dados sobre a possibilidade de um cliente aderir ou não o serviço prestado pela agência com base em testes com múltiplas entradas de dados e com duas saídas possíveis, sim ou não. Seu público alvo são principalmente bancos, qualquer área de estudo sobre comportamento social e estudos sobre aprendizagem de máquina.

III. Metodologia

A partir de uma base de dados que contém informações sobre indivíduos que são clientes de um Banco, fazer o tratamento dos dados, isto é, através de separação entre dados utilizados para os experimentos e dados para o treinamento do classificador e analisar quantitativamente as informações contidas nos campos da base de dados, isto é, fazer uma análise exploratória desses dados. Fazer o treinamento do Classificador Ingênuo de Bayes a partir dos dados, e fazer uma validação. Na próxima etapa usar o classificador para responder a uma pergunta, para cada cliente que está nessa base de dados, "o cliente irá se inscrever em um depósito a prazo?", estes são os experimentos. Analisar os resultados qualitativamente e

quantitativamente, o projeto será dividido nas seguintes etapas:

- Pesquisa sobre o tema: Através da pesquisa bibliográfica, estudar a relevância do assunto, e suas aplicações.
- Base de dados: A base de dados está disponível em [1]. Se refere a uma pesquisa de instituição financeira, possui 45211 instancias e cada instância possui 17 atributos.
- Tratamento dos dados: Fazer a limpeza e seleção dos dados que serão usados no projeto. Os dados selecionados serão divididos em dois grupos, dados para treinamento do classificador e dados para o experimento.
- Analise Exploratória: Através do Uso da biblioteca pandas, numpy e matplotlib, analisar a dispersão do conjunto de dados, através de medidas de tendência central, desvio padrão e o plot de gráficos.
- Classificador Ingênuo de Bayes: O Classificador Ingênuo de Bayes (Naive Bayes) considera em um experimento que as variáveis aleatórias dos campos são todas independentes, mesmo que na realidade elas não sejam, e mesmo assim ele possui uma boa precisão como modelo de classificador.
 - Implementação: Através do uso da biblioteca scikit-learn, numpy e pandas, implementar um classificador ingênuo de Bayes para responder a seguinte pergunta: "o cliente irá se inscrever em um depósito a prazo?"
 - Treinamento do classificador: A partir do conjunto dados previamente separados, treinar o classificador supondo que as vinte variáveis usadas são independentes.
 - Validação: A partir do conjunto de dados restante, verificar se o classificador responde a pergunta corretamente e com que precisão.
- Experimentos: Com o classificador em mãos, realizar alguns experimentos e verificar seus resultados.
- Análise dos resultados: A partir dos dados obtidos nas etapas anteriores, analisar os resultados obtidos.

Cronograma de Atividades

Data	Atividades
02/04/21	Pesquisa Bibliográfica
03/04/21	Pesquisa Bibliográfica
04/04/21	Escrever Resultados no Relatório
05/04/21	Limpeza, Organização e Separação dos dados
06/04/21	Limpeza, Organização e Separação dos dados
07/04/21	Análise Exploratória dos dados
08/04/21	Análise Exploratória dos dados
09/04/21	Análise Exploratória dos dados
10/04/21	Escrever Resultados no Relatório
11/04/21	Implementar o Classificador Ingênuo de Bayes
12/04/21	Implementar o Classificador Ingênuo de Bayes
13/04/21	Implementar o Classificador Ingênuo de Bayes
14/04/21	Treinar o Classificador Ingênuo de Bayes
15/04/21	Consertar Bugs
16/04/21	Validar o Classificador Ingênuo de Bayes
17/04/21	Validar o Classificador Ingênuo de Bayes
18/04/21	Escrever Resultados Obtidos No Relatório
19/04/21	Discussão e Análise dos Resultados Obtidos Durante o Projeto
20/04/21	Produção dos Slides e Gravação do Vídeo
21/04/21	Produção dos Slides e Gravação do Vídeo
22/04/21	Entrega do Projeto

References

- [1] Bank marketing data set. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>