

UC – INTELIGÊNCIA ARTIFICIAL

Título: Projeto Final

Data: 13/04/2023



Grupo de no Máximo de 3 até 5 alunos

Tema: Inteligência Artificial

Objetivo: Orientações Projeto Final

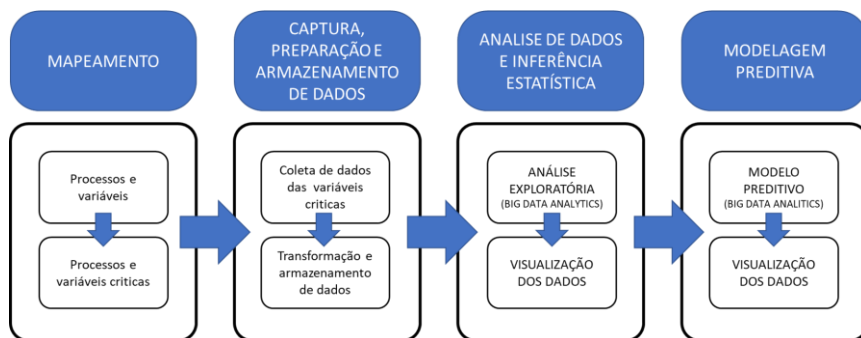
Roteiro: Formar grupos e desenvolver as atividades propostas.

1. Introdução

A computação inteligente é a área da Ciência da Computação que busca a implementação de sistemas computacionais com base no comportamento humano, inspirados na capacidade de solucionar os problemas, bem como aprender com eles (CASTRO, 2009). Os sistemas de Inteligência artificial (IA) têm a capacidade de classificar e prever eventos futuros a partir de um conjunto de dados para treinamento e testes, oferecendo uma probabilidade de ocorrência para cada evento.

É importante que tenhamos nas organizações processos para o desenvolvimento de projetos de IA com etapas bem definidas com métodos de monitoramento e gestão eficientes. A figura 1 apresenta quatro etapas e seus principais objetivos:

Figura 1 - Fases de desenvolvimento de Projetos em I.A



Os problemas ou objetivos tradicionais de pesquisa em IA incluem raciocínio, representação do conhecimento, planejamento, aprendizado, processamento de linguagem natural, percepção e a capacidade de mover e manipular objetos.

Sistemas de IA utilizam métodos estatísticos, inteligência computacional e abordagem simbólica tradicional. Muitas ferramentas são usadas na IA incluindo versões de pesquisa e otimização matemática, redes neurais artificiais e métodos baseados em estatística, probabilidade e economia.

Existem várias etapas que um projeto de IA deve passar até chegar ao aproveitamento pleno da inteligência artificial em startups em estágio de *scale-up* e clientes corporativos.

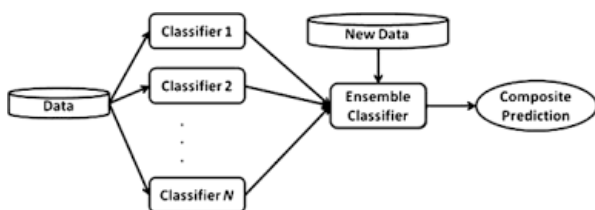
Figura 2 – Hierarquia de necessidades de Projetos de I.A



Quando mais rápido evoluirmos na hierarquia de necessidades da pirâmide, mais rápido o projeto de IA tomará forma e abreviará a entrega de benefícios para a organização. O resultado das análises exploratórias de dados, em cada etapa do ciclo de desenvolvimento como mostra a figura 2, é a melhor compreensão do problema e a busca de formas criativas e inovadoras de resolvê-lo, usando ferramentas apropriadas para o sucesso do projeto.

2. Comitê de Classificadores

O comitê de classificadores (ou ensemble classifier) é um conjunto de classificadores que são utilizados para realizar uma **predição** composta.



Existem alguns motivos para a utilização de um classificador desse tipo:

- Diferentes classificadores baseados em **diferentes aspectos do problema**, por exemplo: classificar uma pessoa por voz, face, escrita...
- **Diferentes grupos de características dentro do mesmo conjunto de dados**, ou seja, algumas instâncias poderiam faltar um atributo, enquanto outras faltaríamos outros atributos. E isso poderia ser tratado com diferentes classificadores, para evitar tratar com apenas um e ter muitos valores faltantes;
- **Diferentes métodos de classificação dependendo do espaço de busca**, por exemplo: quando uma variável tiver textos curtos utilizar um classificador e quando tiver textos longos utilizar outro;
- **Juntar classificadores fracos para formar um forte**, *random forest* é um exemplo disso.

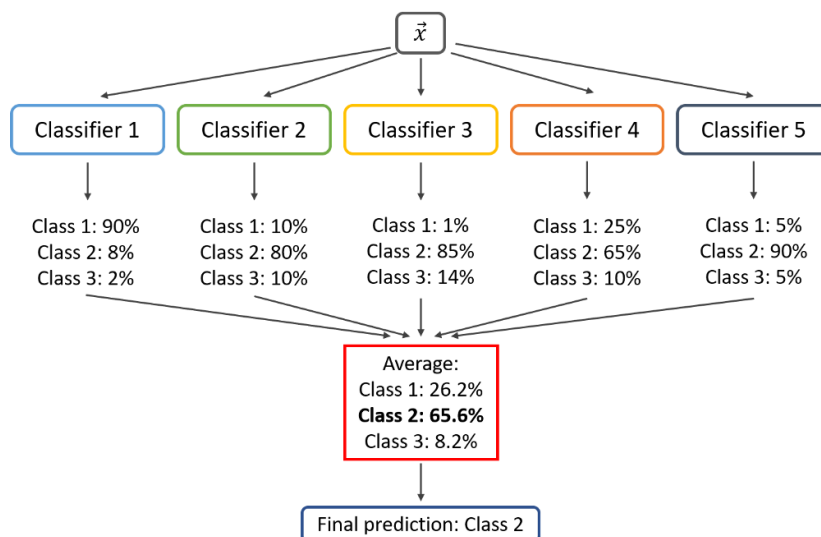
A configuração típica de um comitê possui os seguintes elementos:

- Vários classificadores individuais, que podem variar em: **tipos de algoritmo, características utilizadas e amostras utilizadas**;
- Um combinador, que **escolhe quando cada classificador será chamado** e depois como **combinar os resultados**.

2.1. Arquitetura paralela

Os classificadores são chamados independentes, (Figura3) e os resultados são combinados posteriormente, uma das principais vantagens é a **velocidade**.

Figura 3 - Classificador paralelo



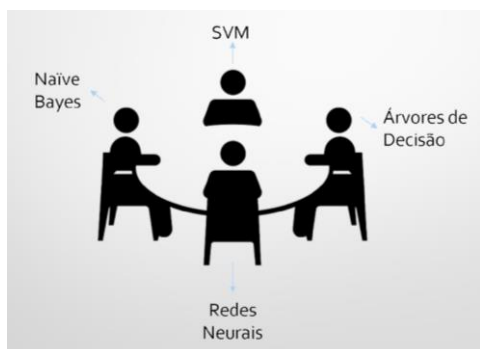
2.2. Classificação Multiclasse

Esse tipo de classificação acontece quando existem mais de 2 tipos de classe no conjunto de dados. Existem vários métodos que podem ser utilizados, alguns deles seriam:

- Regressão Linear
- Técnicas de agrupamento (*clustering*);
- Naive Bayes;
- Árvores de decisão/*random forest*;
- Redes Neurais;
- kNN
- k-Means

3. Proposta do Projeto

O projeto da UC de Inteligência Artificial será fazer um Comitê de Classificadores. Cada grupo analisará **uma base de dados** e aplicará métodos de IA para chegar a algum resultado, após a aplicação desses métodos, deverá compará-los para chegar à conclusão de qual teve o melhor resultado, melhor custo benefício, etc.



3.1. Métodos que deverão ser aplicados

- Árvores de Decisão
- Regressão
- KNN
- Naive Bayes

3.2. Base de dados

Como sugestão poderá ser utilizada a base de dados **Census** ou o grupo poderá escolher uma base de dados entre **10 e 15 variáveis**, conforme o exemplo da base *Census*.

Variáveis da base **Census**

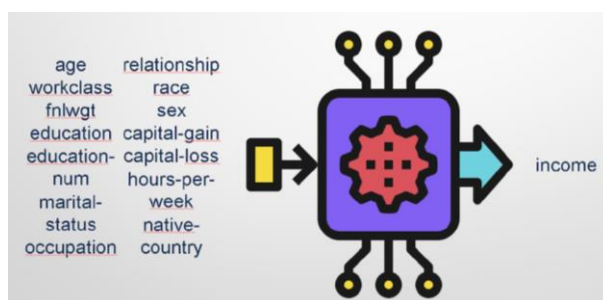
- age
- workclass
- fnlwgt
- education
- education-num
- marital-status
- occupation
- relationship
- race
- sex
- capital-gain
- capital-loss
- hours-per-week
- native-country
- income

3.3. Como chegar aos resultados?

- Estudar a base e compreender todos os parâmetros envolvidos;
- O grupo deverá definir qual será a variável **target** que usará no projeto. Variável **target** será a variável de saída, a variável que se deseja “prever” algo, qual será a conclusão que se quer chegar;
- Aplicar os métodos de IA descritos no **item 3.1**;
- Confeccionar a documentação para cada fase de entrega inicial (**versão1**);
- Comparar os resultados para todos os métodos aplicados
- Confeccionar a documentação final do projeto (**versão final**)

Exemplo (Base de dados Census)

- Podemos inferir a **renda** com base nos demais atributos, por exemplo.



3.4. Cronograma Geral previsto para as Entregas Online

Item	Entregas (online)	Artefatos	Pontos	Data
01	Equipe de Projeto e Base de Dados* (ppt) / (pdf)	Formalização da Equipe e Definição da Proposta do Projeto (* Escolha da Base de Dados)	-----	13/04/2023
02	Explicação da Base de Dados (pdf) / (zip) / (rar)	Os grupos devem escrever sobre a base de dados, explicando as variáveis, o que cada uma representa, quais os tipos de variáveis.	05	27/04/2023 (versão1)
03	Definição da Variável "Target" (pdf) / (zip) / (rar)	Entregar quais as variáveis que precisarão ser transformadas e quais hipóteses podem ser geradas com esta base de dados. Definir a variável target , justificando a escolha dela.	05	27/04/2023 (versão1)
04	Acompanhamento do Projeto (ppt) / (pdf)	<ol style="list-style-type: none"> 1. Árvore de Decisão: Documento contendo os testes e resultados obtidos com a Árvore de Decisão. Esses resultados devem estar justificados. 2. Aplicação de Regressão Linear e Logística: Documento contendo os testes e resultados obtidos com a Regressão. Esses resultados devem estar justificados. 3. Aplicação de KNN: Documento contendo os testes e resultados obtidos com o KNN. Esses resultados devem estar justificados. 4. Aplicação de k-Means: Documento contendo os testes e resultados obtidos com o K-Means. Esses resultados devem estar justificados. 5. Aplicação de Naives Bayes: Documento contendo os testes e resultados obtidos com o Naives Bayes. Esses resultados devem estar justificados. 	10	11/05/2023 (versão1)
05	Apresentação (ppt) e Entrega da Documentação final (Todos os arquivos do Projeto – formato .zip/rar)	<ul style="list-style-type: none"> • Apresentação dos resultados alcançados que devem estar justificados. • Entrega Final do Projeto incluindo os testes. 	10	02/06/2023 (Versão Final)

3.5. Orientações Gerais

O projeto deverá ser feito em grupos de no máximo **3 até 5 alunos**;

- Data da apresentação (Item 1): **20/04/2023 (Tempo de 5 a 10 min.)**
- Entregável: **Formalização da Equipe / *Base de dados .pdf**
- Entregável: Apresentação no formato **.ppt / pptx. (Template Livre)**

ATIVIDADE (A1)	ENTREGÁVEL	ITENS	PONTUAÇÃO
Avaliação Online	Dissertativa (No ULife conforme cronograma)	01	30

ATIVIDADE (A2)	ENTREGÁVEL	ITENS	PONTUAÇÃO
Avaliação Online	Questões fechadas (No ULife conforme cronograma)	01	30

ATIVIDADE (A3)	ENTREGÁVEL	ITENS	PONTUAÇÃO
Projeto Final	(Conforme cronograma de entregas online)	05	30
* Participação e Exercícios em Sala	<ul style="list-style-type: none"> • Apresentações (Arquivo formato .ppt) • Questionários (Arquivo formato .pdf ou zip) 	01	10

TOTAL	100 PONTOS
-------	------------