

**Natasha Gabrielly Porrua**

**Criação de um sistema gestor de base de dados e aplicação de data mining de forma a melhorar gestão e previsão de adoções de gatos na cidade de Bragança**

Dissertação apresentada à Escola Superior Agrária de Bragança para obtenção do Grau de Mestre em Tecnologias da Ciência Animal no âmbito da dupla diplomação com a Universidade Tecnológica Federal do Paraná- Campus Dois Vizinhos

**Orientado por:**

**Professor Doutor Pedro Miguel Lopes Bastos - IPB**

**Professora Doutora Emilyn Midori Maeda - UTFPR**

Bragança

2023

“Nós, seres humanos, estamos na natureza para auxiliar o progresso dos animais, na mesma proporção que os anjos estão para nos auxiliar. Portanto quem chuta ou maltrata um animal é alguém que não aprendeu a amar.”

Chico Xavier

## AGRADECIMENTOS

Este trabalho tornou-se possível através do Programa de Dupla Diplomação existente entre a Universidade Federal Tecnológica do Paraná e ao Instituto Politécnico de Bragança, a primeira em específico o Campus Dois Vizinhos, tem todo o meu carinho e respeito, assim como todo o corpo docente do curso de Zootecnia, que no decorrer da graduação foram excepcionais para a minha formação profissional e crescimento pessoal, o que tornaram esse objetivo possível. E a todos os professores e servidores de ambos os Campus que fazem esse programa acontecer, enriquecendo as vivências acadêmicas e pessoais, abrindo uma porta de oportunidades e realizações.

Ao Diretor da Escola Superior Agrária de Bragança, Professor Doutor Pedro Miguel Lopes Bastos, também orientador desta dissertação, agradeço por todo apoio e dedicação, disponibilidade e conselhos essenciais para a realização deste trabalho.

À Professora Doutora Emilyn Midori Maeda, coorientadora desta dissertação, agradeço por todo suporte e contribuições dadas ao longo do trabalho.

À Ana Isabel Oliveira, que assim como as outras responsáveis pela Associação Amicus Canis, fazem um trabalho excepcional em prol da defesa dos animais, agradeço a disponibilidade e por abrirem as portas da associação para o desenvolvimento deste projeto.

Aos meus pais, Maurício Porrua e Lilian Porrua, por todo amparo e dedicação na minha educação, pelo amor e confiança depositados em mim, que foi essencial para eu chegar até aqui.

Ao meu tio Marcelo Porrua, pelos conselhos e amizade durante esse período, amenizando os efeitos de estar longe da família.

Aos meus irmãos, Mauricio Porrua Junior e Ramon Porrua pela amizade e companheirismo de sempre. Em especial ao Douglas Porrua, por toda motivação, apoio, companheirismo e suporte desde a minha chegada em Bragança até a finalização deste trabalho, foi com certeza muito importante para a minha permanência aqui.

Aos meus amigos que de alguma forma prestaram a sua colaboração, seja presencialmente ou à distância.

Por último, mas não menos importante agradeço a minha namorada Hellen da Silva Campos, que foi definitivamente um pilar essencial para a conclusão desse ciclo. Sou grata por todo amor, apoio e dedicação para que isso se torna-se possível. Por todo cuidado e zelo nos momentos mais difíceis, de incertezas e fragilidade.

A todos que, indiretamente ou diretamente contribuíram para a realização deste trabalho, o meu mais sincero, obrigada.

## RESUMO

Em um cenário comparativo as taxas de adoções infelizmente são significativamente inferiores em relação as taxas de abandono de animais de companhia em Portugal e no mundo. Ressaltando a importância das campanhas de adoção, bem como, o controle populacional das colônias de ruas. O presente trabalho tem como objetivo a criação de uma aplicação informática capaz de auxiliar e otimizar o sistema adotivo, proporcionando a capacidade de desenvolvimento de estratégias mais eficientes, a fim de melhorar as taxas de adoção e controle dos animais. O projeto foi realizado com a Associação Amicus Canis, uma associação localizada na cidade de Braganca – PT, que atua de forma ativa na recolha e controle de animais de rua, majoritariamente os gatos. Os dados foram coletados e registados em *Excel*, posteriormente foi criada a base de dados, realizada com o auxílio do software *Access*, onde se iniciou a criação das interfaces personalizadas e dinâmicas, ou seja, ocorreu a criação de consultas, formulários e relatórios a fim de aprimorar a organização da base de dados, além de facilitar o uso e manutenção da mesma por parte da associação. Utilizando o software *RapidMiner*, foi possível à aplicação da mineração de dados, mais conhecida como *data mining*, responsável por identificar regras e padrões nos dados possibilitando a previsão de resultados. Para isso, utilizou-se uma aprendizagem supervisionada e o algoritmo de Redes Bayes, permitindo a criação do modelo preditivo, obtendo-se 84,78% de performance, ou seja, de possibilidade de adoção. Com o modelo em funcionamento, criaram-se dados reais a três animais fictícios com a intenção de validar o modelo preditivo, tendo como resultado 100% de possibilidade de adoção para os gatos criados de forma fictícia. Considerando a quantidade de dados presentes atualmente na base de dados, justifica-se os resultados obtidos de 100%, ou seja, com o passar do tempo e aumento da base de dados, o modelo irá refinar, tornando-se mais confiável e gerando dados mais condizentes com a realidade. Entretanto, o sistema gestor associado ao modelo predito é funcional e completo, permitindo a previsão de dados, facilitando a tomada de decisão em relação ao sistema adotivo.

**Palavras-chave:** Animais de companhia; Sistema adotivo; Gestão de dados; Mineração de dados.

## **ABSTRACT**

In a comparative scenario, adoption rates are unfortunately significantly lower than the abandonment rates of companion animals in Portugal and around the world. Highlighting the importance of adoption campaigns, as well as population control of street colonies. Therefore, the present work aims to create a software tool capable of assisting and optimizing the adoption system, providing the ability to develop more efficient strategies, in order to improve animal adoption and control rates. The project was carried out with the Amicus Canis Association, an association located in the city of Braganca – PT, which actively works in the collection and control of stray animals, mainly cats. The data was collected and spreadsheeted in Excel, later the database was created, carried out with the help of Access software, where the creation of personalized and dynamic interfaces began, that is, the creation of queries, forms and reports to in order to improve the organization of the database, in addition to facilitating its use and maintenance by the association. Using the RapidMiner software, it was possible to apply data mining for identifying rules and patterns in the data, enabling the prediction of results. For this, supervised activity and the Bayes Network algorithm were used, allowing the creation of the predictive model, obtaining 84.78% performance, that is, the possibility of adoption. With the model in operation, three fictitious animals were created with the intention of validating the predictive model, resulting in a 100% possibility of adoption for the cats created. Considering the amount of data currently present in the database, the results obtained are 100% justified, that is, as time passes and the database increases, the model will refine, becoming more reliable and generating data more consistent with reality. However, the management system associated with the predicted model is functional and complete, allowing data prediction, facilitating decision-making in relation to the adopted system.

**Keywords:** Companion animals; Foster system; Data management; Data mining.

## ÍNDICE

<b>AGRADECIMENTOS.....</b>	<b>III</b>
<b>RESUMO.....</b>	<b>V</b>
<b>ABSTRACT.....</b>	<b>VI</b>
<b>LISTA DE FIGURAS.....</b>	<b>IX</b>
<b>LISTA DE TABELAS.....</b>	<b>X</b>
<b>LISTA DE SIGLAS E ACRÓNIMOS.....</b>	<b>XI</b>
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
<b>2 REVISÃO DE LITERATURA.....</b>	<b>3</b>
2.1 Avanço Tecnológico .....	3
2.2 Base de Dados .....	3
2.2.1 Access .....	4
2.3 Associação Amicus Canis – AMICA.....	4
2.3.1 Gato Doméstico .....	5
2.4 A Descoberta do Conhecimento em Base de Dados .....	7
2.4.1 Data Mining.....	9
<b>3 MATERIAL E MÉTODOS .....</b>	<b>24</b>
3.1 Local de Experimentação .....	24
3.1.1 Controle Populacional.....	24
3.2 Recolha dos Dados .....	24
3.2.1 Excel .....	25
3.3 Access .....	26
3.3.1 Modelo Relacional.....	26
3.3.2 Tabelas.....	26
3.3.3 Consultas .....	28
3.3.4 Formulários .....	30
3.3.5 Relatórios.....	33
3.4 RapidMiner Studio .....	36

3.4.1	<i>Modelo de Predição</i> .....	36
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>41</b>
4.1	Sistema de Gestão de Base de Dados.....	41
4.2	Aplicação do Data Mining .....	41
4.3	Resultados Obtidos .....	41
<b>5</b>	<b>CONCLUSÕES</b> .....	<b>43</b>
5.1	Considerações Finais.....	43
5.2	Recomendação para Trabalho Futuro .....	43
<b>6</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>44</b>

## LISTA DE FIGURAS

<b>Figura 1</b> - Uma visão geral das etapas que compõem o processo KDD.....	8
<b>Figura 2</b> - Fases da metodologia CRISP-DM.....	10
<b>Figura 3</b> - Fases da metodologia SEMMA .....	12
<b>Figura 4</b> - Categorias das tarefas de Data Mining .....	14
<b>Figura 5</b> - Árvore de Decisão .....	17
<b>Figura 6</b> - Rede Bayes .....	19
<b>Figura 7</b> - Modelo do neurônio artificial .....	20
<b>Figura 8</b> - Ilustração do princípio básico da técnica k-NN.....	20
<b>Figura 9</b> - Ilustração da amostra já classificada.....	21
<b>Figura 10</b> - Interface do RapidMiner 10.1.001 .....	22
<b>Figura 11:</b> Planilha com as informações coletadas dos gatos .....	25
<b>Figura 12:</b> Planilha com as informações coletadas dos tutores .....	25
<b>Figura 13</b> - Esquema relacional da BD funcional do sistema .....	26
<b>Figura 14:</b> Consulta dos gatos para adoção em modo de estrutura .....	28
<b>Figura 15:</b> Consulta dos gatos adotados em modo de estrutura .....	28
<b>Figura 16:</b> Consulta por coloração em modo de estrutura .....	29
<b>Figura 17:</b> Consulta de adoções por ano em modo de estrutura.....	29
<b>Figura 18:</b> Formulário dos gatos para adoção – Gato Xavier .....	30
<b>Figura 19:</b> Formulário dos tutores – Tutora Natasha .....	31
<b>Figura 20:</b> Formulário dos gatos adotados – Gato Xavier .....	31
<b>Figura 21:</b> Formulário para consulta de adoções por ano .....	32
<b>Figura 22:</b> Formulário para consulta de animais com a mesma coloração .....	32
<b>Figura 23:</b> Formulário criado como página de arranque .....	33
<b>Figura 24:</b> Processo de criação do modelo preditivo .....	36
<b>Figura 25:</b> Configuração e aplicação do operador validation .....	37
<b>Figura 26:</b> Matriz de Confusão do Modelo .....	38
<b>Figura 27:</b> Teste realizado com o modelo .....	39

## LISTA DE TABELAS

<b>Tabela 1:</b> Estrutura da tabela de registro dos gatos .....	27
<b>Tabela 2:</b> Estrutura da tabela de registro dos tutores.....	27
<b>Tabela 3:</b> Estrutura da tabela de relação.....	27
<b>Tabela 4:</b> Relatório das adoções nos meses de maio e junho .....	34
<b>Tabela 5:</b> Relatório dos animais com coloração amarela .....	34
<b>Tabela 6:</b> Relatório dos animais para adoção .....	35
<b>Tabela 7:</b> Relatório dos animais adotados .....	35
<b>Tabela 8:</b> Resultados do caso de estudo .....	40

## LISTA DE SIGLAS E ACRÓNIMOS

AD	Árvore de Decisão
AMICA	Associação Amicus Canis
BD	Base de Dados
CED	Captura, Esterilização e Devolução
CPADA	Confederação Portuguesa das Associações de Defesa do Ambiente
CRISP-DM	Cross Industry Standard Process for Data Mining
DCBD	Descoberta de Conhecimento em Bases de Dados
DM	Data Mining
GFK	Growth from Knowledge
IR	Indução de Regras
ICNF	Instituto da Conservação da Natureza e das Florestas
KDD	Knowledge Discovery in Databases
k-NN	K-Nearest Neighbours
LPDA	Liga Portuguesa do Direitos do Animal
RB	Redes Bayes
RNA	Redes Neurais Artificiais
RSPCA	World Society for the Prevention of Cruelty to Animals
SEMMA	Sample, Explore, Modify, Model, and Assess
SGBD	Sistema Gestor de Base de Dados
SIAC	Sistema de Informação de Animais de Companhia
XML	Extensible Markup Language

## 1 INTRODUÇÃO

Atualmente, quando se trata de escolha de animais de companhia, os gatos começaram a ser mais considerados. A adoção de gatos encontra-se em expansão com aumento de 78% nos anos de 2020 e 2021 (SIAC, 2021). No entanto, segundo o Instituto da Conservação da Natureza e das Florestas – ICNF (2021), a taxa de animais (cães e gatos) presentes nas ruas também aumenta ultrapassando os 30%. Justamente pela falta de suporte dos meios públicos para realizar o controle, assim como, a falta de investimento em locais de recolha apropriado, a fim de diminuir a reprodução errante que ocasiona a superpopulação dos gatos.

Quanto as adoções, existem algumas características nos animais, que são mais procuradas por futuros tutores, tais como, a cor da pelagem, a idade do animal, tamanho, sexo, personalidade, raça, entre outras. Além de características do próprio adotante, como, a disponibilidade de espaço para o animal, a quantidade de pessoas que vivem na sua habitação e a presença ou não de outros gatos (Stammach e Turner, 1999).

Com isso, surgiu a ideia de associar a tecnologia de mineração dos dados, pois desta forma torna-se possível a identificação de padrões e regras, assim como as correlações perante as características consideradas no momento de adoção. A técnica de Data Mining é bastante famosa e contém resultados incríveis, como, o famoso caso na rede de mercados norte-americana que identificou um padrão nas vendas de fraldas e cervejas nas sextas-feiras, ou seja, através da base de dados da empresa, foi possível descobrir que a maior parte dessas vendas era realizada por homens entre 25 e 35 anos. De forma estratégica a empresa aproximou os displays das fraldas e cervejas na sexta-feira, o que aumentou as vendas de ambos em 30%. (Adaptado de Witten, 2011).

Atualmente as técnicas foram aprimoradas, possibilitando melhores resultados, bem como, a facilidade na interpretação dos padrões. Além de softwares que passaram a serem utilizados, tal como, o Microsoft Access para Microsoft 365 MSO versão 2302, a fim de otimizar a criação do Sistema Gestor de Base de Dados (SGBD), que serve como base para o processo, conhecido como Knowledge Discovery in Databases (KDD), ou seja, a Descoberta de Conhecimento em Bases de Dados (DCBD), realizado na plataforma RapidMiner Studio versão 10.1.001.

Considerando a dificuldade de adoção responsável e controle das colônias, por parte da Associação Amicus Canis (AMICA), uma organização privada sem fins lucrativos, localizada na cidade de Bragança - Portugal. O presente trabalho será focado na identificação

de regras e padrões, quanto as características predominantes e consideradas no momento da adoção. Permitindo a melhor escolha nas estratégias adotivas, possibilitando a rotação dos animais em abrigo e com isso, promovendo o maior controle das populações de rua.

Com isso, espera-se possibilitar maior eficiência quanto as formas de divulgações, maneiras e estratégias utilizadas para as adoções, estimando a probabilidade de os gatos serem adotados, assim como, conhecer quais são os animais com maior potencial adotivo, a fim de otimizar e potencializar os métodos adotivos.

## **2 REVISÃO DE LITERATURA**

### **2.1 Avanço Tecnológico**

Atualmente a tecnologia é a área de maior crescimento, e isso permanece evidente ao longo dos anos, principalmente, as mudanças em todas as tecnologias utilizadas, de acordo com o ramo de estudo. a tecnologia favorece diferentes classes sociais, de diferentes formas. Tal avanço é constante e promissor, sendo o principal ramo de investimento de todas as empresas, nos mais distintos setores, tornando-se na melhor fonte de conhecimento e controle sobre a eficiência empresarial e produtiva.

A par do desenvolvimento tecnológico, encontra-se o desenvolvimento dos meios de armazenamento de informação resultante do desenvolvimento de áreas de aplicação tecnológicas, tal como telemedicina, o ensino à distância, o comércio eletrônico, e afins, tendo como objetivo melhorar e otimizar significativamente a vida da sociedade, assim como elevar a competitividade das empresas (Takahashi, 2000).

A combinação de dois fenômenos resultante do grande fluxo de informação por unidade de tempo, são o aumento significativo da quantidade de informação a tratar e a diminuição do tempo disponível para tal. O primeiro fenômeno acima identificado, é explicado pela necessidade de obtenção de conhecimento potenciado pelo aumento da capacidade de armazenamento da informação. O segundo fenômeno, apoiado igualmente no avanço tecnológico, especificamente na evolução dos sistemas informáticos capazes de velocidades de processamento cada vez maiores, permitindo a troca de informações em tempo real em qualquer lugar do mundo, onde a distância deixa de ser um obstáculo (Le Coadic, 1996).

### **2.2 Base de Dados**

O armazenamento de informação resulta numa base de dados (BD), seja ela, uma lista numa folha de papel ou até mesmo uma folha num programa informático. Uma BD é toda e qualquer maneira de armazenar de forma organizada a informação, a fim de facilitar a sua pesquisa ou consulta.

À medida que a quantidade de dados aumenta, a probabilidade de incoerência, dados repetidos ou errôneos também aumenta. Com isso, cresce a necessidade de automatização, ou seja, o uso de ferramentas com especialidade em armazenar e criar uma base de dados.

### **2.2.1 Access**

Considerado um Sistema de Gestão e Base de Dados – SGBD, tendo como objetivo o registro e manutenção de informação considerada necessária à organização/utilizador final. É um sistema com possibilidade de criação e manipulação de bases de dados.

O Access é um programa de Base de Dados Relacional, projetado exclusivamente para ser um aplicativo do Windows, ganha espaço no mercado principalmente pela qualidade das ferramentas oferecidas, assim como, por ser um sistema relacional, ou seja, conta com a existência de relações entre as entidades/tabelas, conectando as mesmas, promovendo maior otimização e qualidade para a BD.

Os principais elementos que pertencem ao BD do Access, são as tabelas, formulários, relatórios, consultas, macros e módulos. Estes, promovem a eficiência na organização dos dados, na visualização e interpretação do utilizador final.

### **2.3 Associação Amicus Canis – AMICA**

Fundada em 2011, trata-se de uma organização privada sem fins lucrativos, gerida por uma equipe principal de voluntárias responsáveis pela organização geral da associação, onde as tarefas de limpeza, cuidados e capturas dos animais são realizadas por elas, bem como, por estagiários que participam como voluntários.

A sede encontra-se na cidade de Bragança, atua no combate contra o abandono e maus tratos em prol das causas animais, majoritariamente cães e gatos, tendo como maior atuação o cuidado com as colônias felinas presentes na cidade, objetivando o controle da população e adoção dos mesmos. A organização opera nas colônias capturando os animais para realizar a esterilização e desparasitação, sendo realizadas por clínicas veterinárias privadas mediante parceria com a AMICA, além de contar com cuidadores responsáveis por alimentar e zelar pelos animais das ruas. Os animais com potencial adotivo são recolhidos de acordo com a disponibilidade de espaço.

Atualmente o método de controle dos animais adotados é dado através do preenchimento de formulários físicos, com informações pessoais do tutor e do animal a ser adotado, ou seja, a fonte de base de dados atual, consiste em ficheiros impressos e anexados a uma pasta. Dado o passar dos anos, muitos arquivos foram perdidos ou até mesmo danificados por desgaste, além de tal método dificultar a acessibilidade e busca por informação específica.

### **2.3.1 Gato Doméstico**

A trajetória dos gatos domésticos ao longo dos anos e em diferentes regiões, teve altos e baixos, os animais foram de venerados para perseguidos, permanecendo durante muito tempo essa oscilação a respeito dos gatos e sua representação (Serpell, 2000). Isso certamente reflete nos dias de hoje, não só no comportamento do felino, mas como na sua relação com humanos.

A domesticação dos gatos carece de algum estudo, no entanto, segundo Sparkes (2016), “sabe-se que há dez mil anos ainda não existiam gatos domésticos como os conhecemos, tendo início muito depois da domesticação do cão”. Logo, os gatos permanecem com muitos comportamentos instintivos do seu estado selvagem, sendo considerados independentes, solitários e com capacidade de viver sem tutor, justamente com terem sido os únicos animais que se auto domesticaram, ou seja, se aproximaram dos locais urbanos a procura de alimentos e abrigos. De forma que essa aproximação foi ganhando força e espaço.

Concebe-se que o número de gatos como animais de companhia cresce ao redor do mundo (Rochlitz et al., 2000), com população de 113 milhões somente na Europa, os gatos representam 24% dos animais domésticos em lares europeus (European Pet Food Industry Federation, 2021).

#### **2.3.1.1 Sistema de Adoção versus Abandono**

Considerando a mudança no estilo de vida da sociedade, ou seja, com o aumento da urbanização, majoritariamente as pessoas passaram a viver em apartamentos, optando pela redução na quantidade de filhos, deste modo, às adoções de gatos tornaram-se uma opção mais viável, devido à alta adaptabilidade, a menor exigência quanto ao espaço para viver, a limpeza, a independência e interação dos felinos (Faraco, 2013).

Em Portugal, segundo um estudo conduzido pela Growth from Knowledge (GFK) (2015), embora o número de canídeos ainda seja superior, a adoção de felinos pelos portugueses aumentou para 25% face aos 3% no ano 2011, tendo havido também um aumento do número de gatos por habitação. Nos últimos anos, segundo o Sistema de Informação de Animais de Companhia (SIAC, 2021), a adoção de gatos teve um aumento de 78%, ou seja, 629.914 animais estão registrados no novo sistema SIAC. Esse aumento originou-se devido a pandemia, a maior disponibilidade e solidão das pessoas abriram espaço para adoções de animais de companhia, uma vez que os animais têm uma função essencial para a estabilidade emocional das pessoas e isso foi devidamente reconhecido por parte das mesmas.

### 2.3.1.1.1 Comportamento dos Gatos

O comportamento dos felinos são fatores decisivos e atraem os tutores para as adoções, justamente por não necessitarem de atenção o tempo todo, por serem mais independentes. Assim como, o hábito alimentar fracionado, ou até mesmo por suportar distâncias maiores entre as refeições. O hábito da limpeza também é bem-visto, serem mais higiênicos e terem o costume de urinar e defecar nas caixas de areia, o que facilita a organização e limpeza do ambiente.

Em contrapartida, certos comportamentos estão entre os fatores que podem levar ao abandono. Logo, a comparação errônea entre cães e gatos eleva as expectativas dos tutores e quando os felinos não correspondem, ocorre a desconexão na relação.

Com isso, comportamentos agonísticos, de independência, marcação de território ao arranhar superfícies, reatividade, escalada e saídas da residência, frustram os tutores podendo levar ao abandono do animal (Machado et al., 2017). No entanto, como já dizia Serpell (1996), “é importante enfatizar que muitos comportamentos entendidos como problemáticos são, na verdade, naturais. O que ocorre é a desconexão entre o que o tutor espera do seu animal e o que ele de fato exhibe.”

Podberscek (1997), enfatiza a importância da educação, não só para gerar mais respeito pelos animais em geral, mas também para reduzir a incidência deste tipo de abuso devido à ignorância.

### 2.3.1.2 Legislação Portuguesa

Utilizando como referência a LPDA – Liga Portuguesa do Direitos do Animal, fundada em 1981, tratando-se de uma associação de utilidade pública, sem fins lucrativos e de âmbito nacional. É associada da World Society for the Prevention of Cruelty to Animals (RSPCA), está registrada na Agência Portuguesa do Ambiente com o nº 22N e é sócia fundadora da Confederação Portuguesa das Associações de Defesa do Ambiente (CPADA).

- **Convenção Europeia para a Proteção dos Animais de Companhia:** aprovada para ratificação em 13 de abril de 1993, aberta à assinatura dos estados membros do Conselho da Europa em 13 de novembro de 1987. Onde, se reconhece “Que o homem tem uma obrigação moral de respeitar todas as criaturas vivas”.

- **Estatuto Jurídico dos Animais:** lei 8/2017 de 3 de março, estabelece um estatuto jurídico dos animais, reconhecendo que são seres vivos dotados de sensibilidade e objeto de proteção jurídica em virtude da sua natureza.
- **Classificações e Definições:** decreto-lei 314/03, de 17 dezembro define o animal de companhia como, qualquer animal possuído ou destinado a ser possuído pelo homem, designadamente em sua casa, para seu entretenimento e enquanto companhia.
- **Vacinação:** é obrigatória a vacina antirrábica dos cães com mais de 3 meses de idade, atualizada anualmente. A vacinação antirrábica de gatos é recomendada, mas voluntária. (Portaria n.º 264/2013, de 16 de agosto).
- **Chip de identificação electrónica:** os cães e os gatos devem ser identificados por método electrónico e registados entre os 3 e os 6 meses de idade. (Decreto-lei 313/03, de 17 dezembro).
- **Maus tratos a animais:** São proibidas todas as violências injustificadas contra animais, consistentes em, sem necessidade, infligir a morte, o sofrimento cruel e prolongado ou graves lesões a um animal. (Lei 92/95 de 12 setembro).

Sob punição com pena de prisão até um ano ou com pena de multa até 120 dias. Se dos fatos previstos resultar a morte do animal, a privação de importante órgão ou membro ou a afetação grave e permanente da sua capacidade de locomoção, o agente é punido com pena de prisão até dois anos ou com pena de multa até 240 dias. (Lei 69/14 de 29 agosto e Lei 110/15 de 26 agosto).

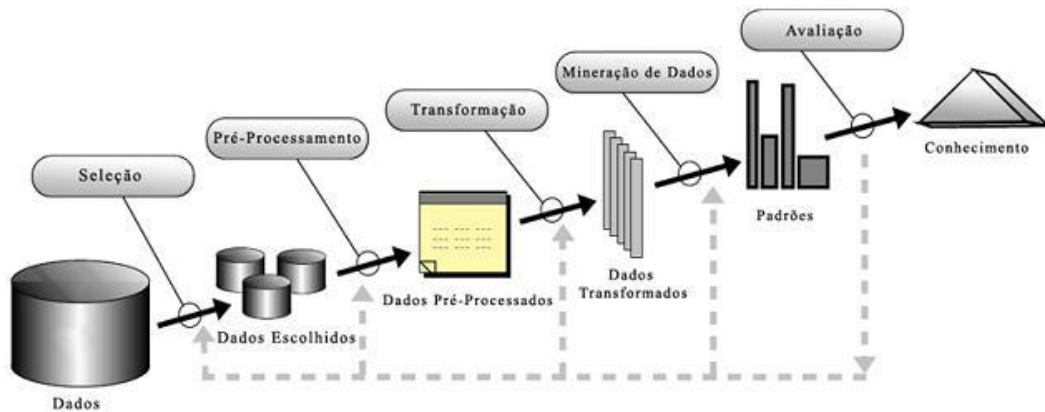
- **Abandono:** É proibido abandonar intencionalmente na via pública animais que tenham sido mantidos sob cuidado e proteção humanas, num ambiente doméstico ou numa instalação comercial ou industrial. (Lei 92/95).

Sob punição com pena de prisão até seis meses ou com pena de multa até 60 dias. (Lei 69/14).

## 2.4 A Descoberta do Conhecimento em Base de Dados

Tendo em vista o avanço tecnológico, assim com o aumento na capacidade de armazenamento de dados em grande escala, tornou-se primordial o desenvolvimento de meios e formas de processar e utilizar tais dados armazenados, a fim de torná-los informação útil. Assim, no final dos anos 80, surgiu o processo KDD – Knowledge Discovery in Databases, ou seja, a Descoberta de Conhecimento em Bases de Dados – DCBD (Apte, 2002).

Segundo Fayyad et al., (1996), a definição do processo de KDD é demonstrada como “a extração não trivial de informação implícita, previamente desconhecida e potencialmente útil, a partir dos dados”, sendo um processo iterativo e iterativo, contendo diversas fases, representadas na figura 1.



**Figura 1** - Uma visão geral das etapas que compõem o processo KDD

**Fonte:** Fayyad et al., 1996

- **Seleção dos Dados:** propriamente dito, consiste na fase em que serão selecionados os dados relevantes dentre todos os existentes na base de dados. São essas informações que farão parte do processo de KDD. Podem ser classificadas como internas, sendo os dados que já estão incorporados no sistema da aplicação utilizada. E as externas, estão presentes fora do sistema, necessitando de formatação e seguidamente serem inseridos nos sistemas, a fim de poderem ser recuperados e utilizados (Bastos, 2015).
- **Pré-processamento dos Dados:** representa a limpeza dos dados, correção de ruídos, dados errôneos e inconstantes. Garante a qualidade, veracidade e integridade dos fatos representados (Boente et al., 2008)
- **Enriquecimento e Transformação dos Dados:** expressa-se na forma como os dados serão codificados, garantindo uniformidade, normalização e coerência para que possam ser usados plenamente pelos algoritmos de mineração de dados (Boente et al., 2008).
- **Mineração dos Dados ou Data Mining:** é a principal etapa do processo de KDD, onde ocorre a busca por informações úteis, através das técnicas de Inteligência Computacional de acordo com o objetivo acordado, que traçam similaridade ou discordâncias entre os dados, a fim de encontrar padrões e/ou regras, potencializando as informações para serem utilizadas nas tomadas de decisões (Dantas et al., 2008).

- **Interpretação dos Dados:** nessa fase ocorre a leitura dos novos padrões, podendo ser realizada com o auxílio de ferramentas estatísticas e de visualização que possibilitam a verificação da validade das informações. Caso não ocorra como o esperado, pode-se retornar nas etapas anteriores, a fim de melhorá-las para atingir resultados válidos (Bastos, 2015).

### 2.4.1 Data Mining

Sendo a principal etapa do processo de KDD, o Data Mining ganhou força, justamente pela necessidade de processar grandes quantidades de dados armazenados. Muitas informações eram passadas despercebidas pela falta de tecnologia capaz de processar e identificar conhecimento presente de forma oculta nas bases de dados.

Desta forma, o KDD e DM, estão em constante avanço para atender a demanda exigida no mercado, visto que a informação é o ativo mais importante para os negócios das organizações, sendo essencial para o aumento da competitividade de grandes, médias e pequenas empresas. Estando atrelado diretamente com as tomadas de decisões a fim de melhorar a eficiência da empresa, nos mais diferentes ramos (Dantas et al., 2008).

São muitas as definições para o DM, no entanto, sempre mantendo o contexto, segundo Carvalho (2005), é “definimos *datamining* como o uso de técnicas automáticas de exploração de grandes massas de dados de forma a descobrir novos padrões e relações que devido ao volume de dados, não seriam facilmente descobertas a olho nu pelo ser humano.”

Dentre as motivações para o uso do DM, encontra-se principalmente a grande capacidade e disponibilidade de dados armazenados eletronicamente, que podem ser utilizados como informação auxiliando na previsão de um conhecimento futuro, segundo Dantas et al., (2008), “descobrir informações sem uma prévia formulação de hipóteses e buscar por algo não intuitivo, é na verdade tornar dados sem obviedade em valiosas informações estratégicas.”

#### 2.4.1.1 Metodologias Utilizadas

Priorizando a compreensão, implementação e desenvolvimento do processo de DCBD torna-se mais fácil se for enquadrado no contexto de uma metodologia (Adapt. Bastos, 2015)

Segundo Nogueira (2014), “O CRISP-DM é uma metodologia genérica que consiste numa sequência de passos usualmente envolvidos num projeto de data mining. A outra, o

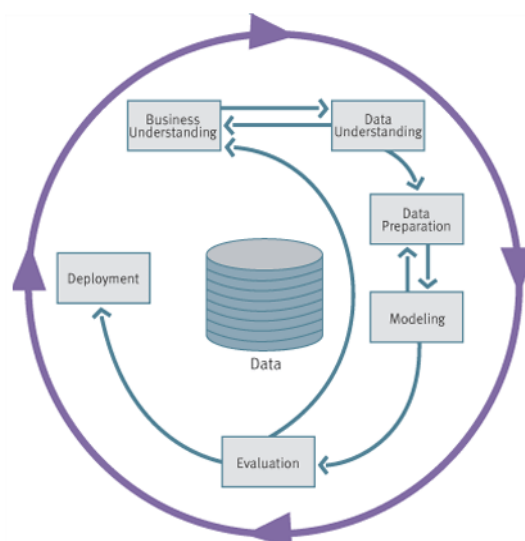
SEMMA, é uma metodologia específica para a ferramenta SAS Enterprise Miner. Embora não sejam necessárias todas as fases de cada metodologia para todos os projetos, estas metodologias, estabelecem uma base para a estratégia a adotar para o correto desenvolvimento de um projeto de data mining.”

#### 2.4.1.1.1 Metodologia CRISP-DM

O CRISP-DM é um acrônimo para Cross – Industry Standard Process for Data Mining que pode ser traduzido como Processo Padrão Inter-Indústrias para Mineração de Dados. O CRISP-DM é uma metodologia genérica para implementação de projetos de DM e ajuda na resolução de problemas típicos de DM (Chapman et al., 2000).

Em 1996, quando a DM ainda estava no início do seu desenvolvimento, no entanto emanava potencialidade, surgiu a necessidade de desenvolver uma metodologia definida e documentada para a implantação da mineração de dados, de forma que fosse gratuita, independente da indústria, sem proprietário e que tivesse capacidade de auxiliar as empresas no desenvolvimento dos projetos com DM. Logo, quatro figuras fortes na área do DM, sendo elas, a Daimler-Benz, a Integral Solutions Ltd, a NCR e a OHRA. Criaram o que veio ser chamado de CRISP-DM. No ano seguinte ocorreu um consórcio nomeado de CRISP-DM SIG, onde ocorreram muitos ajustes com a colaboração de mais profissionais na área, lançando quatro anos depois a versão 1.0 do CRISP-DM. Permanecendo sem atualizações desde então (Adapt. Gomes, 2011).

A metodologia conta com 6 fases no seu ciclo de vida, demonstradas na figura abaixo:



**Figura 2** - Fases da metodologia CRISP-DM

**Fonte:** Adapt. Por Amorim, 2006

O ciclo de vida da metodologia pode ser descrito de acordo com (Chapman et al., 2000):

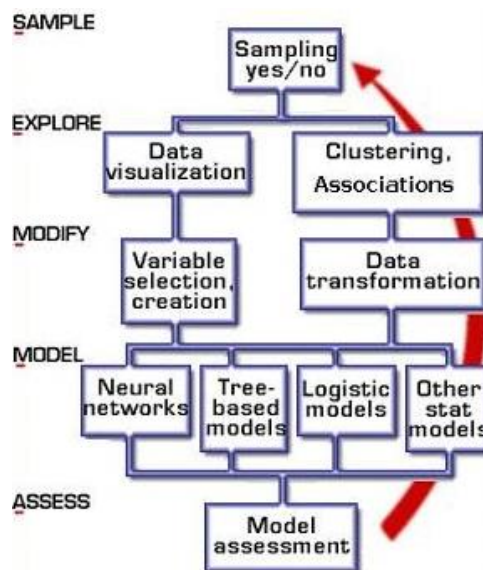
- **Estudo do Negócio (Business Understanding):** possivelmente trata-se da fase mais importante, pois consiste na validação no processo para tal problemática. Com tudo, deve ocorrer a identificação de um problema a ser solucionado, juntamente dos objetivos e requisitos do projeto, e então, a conversão do conhecimento avaliando a necessidade de implementação da mineração de dados no determinado projeto/organização.
- **Exploração dos Dados (Data Understanding):** consiste no primeiro contato da equipe com os dados a serem utilizados, visando a familiarização entre ambos. De modo que ocorre uma análise descritiva das características das variáveis, descartando possíveis erros e estabelecendo o primeiro conjunto de hipóteses, assim como, identificar possíveis conjuntos com potencial para utilizar no estudo.
- **Preparação dos Dados (Data Preparation):** conta com a soma das tarefas e ferramentas para preparar o conjunto de dados de forma que estes supram os requisitos da próxima etapa, a modelação. Portanto, nesta fase ocorre a seleção, limpeza e transformação dos dados com ruído, ou inconsistentes e estranhos, visando a construção de um conjunto de dados final em condições para a próxima etapa.
- **Modelação dos Dados (Modeling):** nesta fase, existem dois tipos básicos de modelos, aqueles que classificam e aqueles que preveem. Portanto tais técnicas de modelagem são aplicadas nos dados a fim de obter valores otimizados, através da aplicação de algoritmos de forma a procurar, identificar e exibir quaisquer padrões relativamente aos dados. De forma geral, as técnicas atuam nos mesmos tipos de problemas na mineração, mas pode ocorrer de uma técnica exigir algum requerimento específico na forma dos dados, neste caso, voltar para a etapa de preparação torna-se necessário. Aliás, tal ação é bastante comum quando se trata do processo de DM.
- **Avaliação do processo (Evaluation):** nesta etapa ocorre a avaliação criteriosa do processo, com medidas de qualidades adequadas, assim como revisão de todos os passos tomados ao longo do processo. Certificando-se que o modelo criado irá suprir a demanda, bem como os objetivos colocados ao processo de DM. Com essa revisão detalhada, ao fim deve-se tomar uma decisão sobre a viabilidade dos resultados da mineração.

- **Execução (Deployment):** esta etapa muitas vezes é realizada pelo cliente, considerando boas soluções, os modelos obtidos são incorporados em processos de tomada de decisão e em novos estudos analíticos. Os resultados podem ser repassados em formato de relatório para o cliente, ou até mesmo executar processos de mineração de dados repetidamente, visando o entendimento do cliente para então ocorrer a integração dos resultados nas atividades da organização.

#### 2.4.1.1.2 Metodologia SEMMA

Esta metodologia corresponde a um acrônimo de cinco fases do processo (figura 4), sendo elas, Sample (Amostragem), Explore (Exploração), Modify (Modificação), Model (Modelação) e Assessment (Avaliação). Consiste na aplicação das ferramentas do SAS Enterprise Miner nas tarefas de um projeto de DM, devido a facilidade de entendimento, otimiza a organização, manutenção e visualização dos projetos de mineração. Contribuindo através da exploração estatística, selecionando e transformando os valores preditivos com potencial de uso, modelando os mesmos para prever os resultados, promovendo uma melhor eficiência do processo, a fim de confirmar a precisão do modelo utilizado (Olson & Delen, 2008).

O SEMMA também funciona como um ciclo de experimentação altamente iterativo, portanto ao deparar-se com resultados não satisfatórios, existe sempre a opção de retornar nas etapas anteriores e ajustar as variáveis, obtendo um modelo mais eficaz (Olson & Delen, 2008).



**Figura 3** - Fases da metodologia SEMMA

**Fonte:** Olson & Delen, 2008

As fases podem ser descritas e detalhadas de acordo com (Olson & Delen, 2008):

- **Amostragem (Sample):** realiza-se a retirada de uma porção do conjunto de dados estudados, a amostra deve ser suficiente para ter boa representatividade, em contrapartida, não deve comprometer a facilidade e rapidez do processamento. Reduzindo os custos e desempenho computacional a técnica de amostragem é defendida e utilizada amplamente. Tratando-se de bancos de dados de grande escala, aconselha-se a mineração de uma parcela representativa do volume todo, promovendo a redução drástica do tempo de processamento para obter as informações importantes para o estudo.
- **Exploração (Explore):** ocorre a exploração de forma visual ou numérica, com a utilização de gráficos, tabelas, histogramas, entre outros, a fim de identificar parâmetros semelhantes, anomalias, tendências, auxiliando na percepção de possíveis grupos. Com isso, o processo de descoberta é facilitado, uma vez que os dados estão refinados e redirecionados. Buscando uma maior eficiência pode-se usar técnicas mais avançadas.
- **Modificação (Modify):** a fim de tornar as variáveis mais significativas, ocorre a transformação dos dados, seleção e inclusão de novas informações, assim com a redução das variáveis.
- **Modelação (Model):** nesta etapa ocorre a criação de modelos com base nas técnicas de aprendizagem automática e estatística, ou seja, serão selecionados os modelos que correspondem com o universo estudado, bem como, que suprem a demanda e os objetivos propostos pelo estudo.
- **Avaliação (Assessment):** consiste na aplicação do modelo, onde espera-se a confiabilidade dos dados, podendo ser aplicado em um conjunto de teste, podendo assim ocorrer alguns ajustes e reparos, caso se faça necessário.

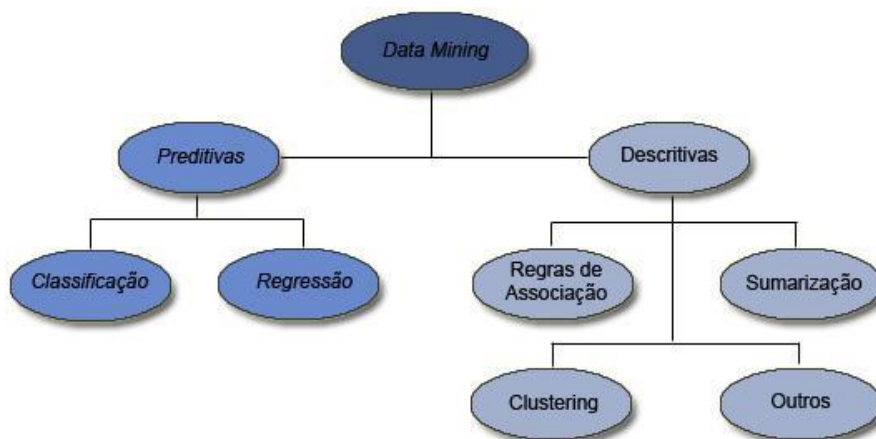
Visando a resolução de problemas de forma ágil e rápida, ambas as metodologias correspondem, no entanto, em um cenário comparativo e em âmbito geral, a CRISP-DM, aparenta ser mais completa do que a SEMMA, justificando a sua maior utilização, bem como, devido ao fato desta metodologia ser aplicada a qualquer ferramenta. Diferentemente da SEMMA, tendo aplicação em uma ferramenta específica.

### 2.4.1.2 Objetivos do Data Mining

De acordo com o que se espera do processo de DCBD, existem dois conceitos importantes a serem definidos, sendo a tarefa e a técnica utilizada. É crucial saber a diferença entre ambas, para garantir o êxito nos resultados, com isso devem ser definidas com base no tipo de informação usado, bem como no conhecimento almejado (Cavalcante, 2014).

Segundo Amo (2004), a tarefa consiste na especificação (objetivos) do que pretendemos encontrar nos dados, que tipo de regularidades ou categoria de padrões temos interesse em encontrar, ou que tipo de padrões nos poderiam surpreender. A técnica de DM consiste na especificação de métodos (algoritmos) que nos garantam como descobrir os padrões que nos interessam.

Na DM há diferentes objetivos e algoritmos que podem ser utilizados, os objetivos geralmente enquadram-se em duas categorias principais, sendo de previsão e descrição.



**Figura 4** - Categorias das tarefas de Data Mining

Fonte: Adapt. Domingues, 2004

#### 2.4.1.2.1 Tarefas Preditivas

Basicamente os valores futuros, ou seja, valores previstos, são valores estimados através da criação de um modelo de previsão utilizando os valores já conhecidos. Logo, a previsão estima valores futuros de variáveis de interesse, com base em algumas variáveis, bem como possibilita a descoberta de padrões de comportamento dessas mesmas variáveis a partir de exemplos, automatizando e facilitando a tomada de decisão (Bastos, 2015).

#### **2.4.1.2.1.1 Classificação**

Trata-se da criação de classes pré-definidas, com atributos previamente escolhidos, onde os algoritmos procuram relações entre os dados os classificando de acordo com suas características. Essas informações são cruzadas com as características das classes, ocorrendo a identificação da classe em que tais dados pertencem (Cavalcante, 2014).

#### **2.4.1.2.1.2 Regressão**

É uma técnica baseada na previsão, uma vez que se propõem a previsão de valores (numéricos) futuros baseando-se em uma variável dependente, através de exemplos. O objetivo corresponde à concepção de um modelo capaz de mimetizar uma função desconhecida que se aproxime da função dada por um conjunto de vetores etiquetados. (Quintela et al., 2005). Para Amorim (2006), “a arte de estimar é exatamente esta: determinar da melhor forma possível um valor, baseando-se em outros valores de situações semelhantes.”

#### **2.4.1.2.2 Tarefas Descritivas**

Basicamente trata-se da utilização dos próprios dados em que se está trabalhando, buscando entender e destrinchar os detalhes, a fim de coletar a maior quantidade de informação possível. Interpretando e reconhecendo os padrões, grupos, anomalias, correlações etc., logo, consiste em um modelo exploratório.

#### **2.4.1.2.2.1 Associação**

Consiste em uma técnica que se compara as correlações e dependências entre as variáveis, quando uma variável majoritariamente depende de outra para aparecer, entende-se como regra de associação, e então facilita a tomada de decisão perante tal situação.

#### **2.4.1.2.2.2 Sumarização**

Essa técnica está atrelada a descrever características específicas e comuns entre variáveis presentes em um mesmo grupo, a fim de identificar padrões melhorando as estratégias de marketing de uma empresa.

#### 2.4.1.2.2.3 Segmentação ou Clusterização

É a tarefa onde ocorre a criação de classes de acordo com a similaridade das variáveis, diferente da associação onde as características já estão pré-definidas, na segmentação as classes são criadas após o cruzamento das variáveis. O objetivo principal é a similaridade inter-segmentos ser baixa e a similaridade intra-segmentos ser alta, ou seja, as variáveis semelhantes se aproximam criando um cluster, e proporcionalmente se distanciam das variáveis diferentes, onde essas também criam um cluster (Cavalcante, 2014).

#### 2.4.1.2.2.4 Visualização

Trata-se da melhor visualização das informações, ou seja, dinamizar os resultados para facilitar o entendimento, geralmente utilizando gráficos dinâmicos e de fácil compreensão (Bastos, 2015).

#### 2.4.1.3 Técnicas para o Data Mining

Deve destacar-se que não existe um modelo universal de DM que resolva, de forma eficiente, todos os problemas. A escolha de um determinado algoritmo de inteligência artificial é de certa forma uma arte (Fayyad et al., 1996), uma vez que existem diferentes modelos para as mesmas tarefas de DM com vantagens e desvantagens intrínsecas (Bastos, 2015).

E atualmente existe um vasto leque de algoritmos de inteligência artificial disponíveis para a implementação de modelos preditivos, no entanto a escolha costuma recair num pequeno conjunto de técnicas, que são reconhecidas como sendo mais eficazes e apresentam a melhor relação entre simplicidade de implementação e eficácia (Gomes, 2011).

As técnicas contam com um tipo de aprendizagem, ou seja, a forma como o algoritmo aprende e executa sua funcionalidade. Portanto, existem duas formas de aprendizagem.

- **Atividade Supervisionada:** que consiste na utilização de dados para treino e dados para a validação, neste tipo, ocorre o envolvimento de um “professor” para fornecimento de respostas corretas à rede. Existindo uma vaga ideia do que se está procurando, onde as respostas produzidas pelo algoritmo são comparadas com as corretas (Côrtes et al., 2002) e (Bastos, 2015).
- **Atividade Não Supervisionada:** não existindo uma meta definida, o algoritmo atua caracterizando e segmentando os dados, a fim de descobrir alguma estrutura

significativa. Traçando padrões sem ter fornecimento externo em relação ao resultado correto (Côrtes et al., 2002) e (Bastos, 2015).

Com isso, as técnicas são caracterizadas dentro de um tipo de aprendizagem, podendo existir várias opções de algoritmos de inteligência artificial, no entanto serão descritas a seguir apenas os principais utilizados.

#### 2.4.1.3.1 Árvore de Decisão

Sendo o algoritmo mais utilizado no data mining, a árvore de decisão utiliza uma representação gráfica que facilita a compreensão dos resultados obtidos. Segundo Quinlan (1986), “numa árvore de decisão, cada folha (nodo terminal) representa uma classe e cada nodo intermediário representa um teste, envolvendo uma ou mais variáveis/atributos e cada possível resultado do teste origina uma nova subárvore”.



**Figura 5** - Árvore de Decisão

**Fonte:** Didática Tech

Como podemos ver, todas as vezes em que a coluna (primeiro nó) “Sol?” é igual a “Não”, a nossa árvore de decisão responderá que a pessoa não foi para praia. Porém quando esta resposta for “Sim”, temos casos em que a pessoa foi para praia e outros em que ela não foi. Ou seja, é necessário fazermos mais uma pergunta para definirmos a resposta, levando assim ao segundo nó de nossa árvore, que irá conferir a informação existente na coluna “Vento?”.

Para a criação e definição da árvore existem duas métricas que podem ser utilizadas (Quinlan, 1998).

- **Entropia:** o algoritmo confere o ganho de informação de cada variável. Aquela que apresentar maior ganho de informação será a variável do primeiro nó da árvore.
- **Índice GINI:** é calculado para cada atributo, utilizando para o nó o atributo com menor índice, isto é, o atributo mais informativo é aquele que tem menor índice de impureza.

#### 2.4.1.3.2 Indução de Regras

A indução de regras (IR) está diretamente associada a uma árvore de decisão (AD), uma vez que através da AD pode ocorrer a extração de regras/tendências representadas pelo modelo. Com isso, o principal objetivo é identificar dependências entre os valores, ou seja, se um determinado atributo está presente, certamente outro atributo também está presente (Toloo et al., 2009). Essas regras contam com dois parâmetros, sendo:

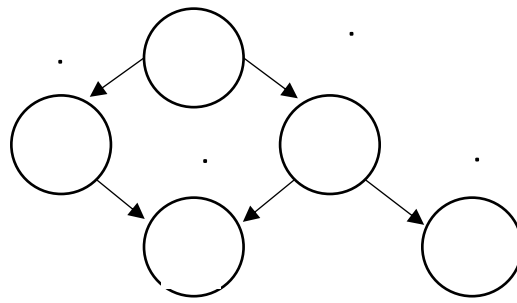
- **Grau de Confiança:** consiste na probabilidade de a regra verificar-se.
- **Nível de Suporte:** trata-se do número de casos em que a regra se verifica.

As regras devem atender certas propriedades, como, previsões assertivas, ou seja, se maioritariamente o SE da regra for verdadeiro, o ENTÃO também é. Assim como, serem de fácil compreensão e serem úteis para a tomada de decisão do utilizador.

As principais vantagens deste algoritmo são, o modo direto de lidar com os dados, desempenho, explicação e compreensão das regras, e os detalhes da solução do problema sem facilmente observáveis (Quinlan, 1998).

#### 2.4.1.3.3 Redes Bayes

Técnica que se baseia no Teorema de Bayes, logo, consiste na detecção numérica da probabilidade que um evento pode ocorrer, baseada na probabilidade de um evento que já ocorreu (Camilo, 2009). As RB podem ser consideradas diagramas que organizam o conhecimento através de um mapeamento entre causas e efeitos. Sendo constituída por estrutura gráfica e parâmetros numéricos, às RB apresentam um alto nível de eficiência na análise de dados com relações de precessão temporal, podendo gerar automaticamente predições ou decisões mesmo na situação de inexistência de alguns dados utilizados na modelação (Heckerman et al., 1995).



**Figura 6** - Rede Bayes

**Fonte:** Bastos, 2015

Segundo Lampinen (1998), “Para a representação de conhecimento, as RB consistem em grafos acíclicos com as seguintes características: os nós correspondem a variáveis aleatórias e uma ligação direcionada ou arco liga pares de variáveis. O significado de um arco dirigido do nó Z ao nó Y é que Z tem uma influência direta sobre Y. Cada nó tem associado os estados da variável e uma tabela de probabilidades condicionadas que quantifica os efeitos que os pais exercem sobre um nó (probabilidade de o nó estar num estado específico dado os estados dos seus pais)”.

De acordo com Zhang (2004), “comparativos mostram que os algoritmos Bayesianos, chamados de *naive Bayes*, obtiveram resultados compatíveis com os métodos de árvore de decisão e redes neurais. Devido a sua simplicidade e o alto poder preditivo, é um dos algoritmos mais utilizados”.

A designação “naive” em Naive Bayes é-lhe atribuída pelo facto de o algoritmo usar técnicas Bayesianas, mas não considerar dependências que possam existir. Portanto, as suas suposições são consideradas ingênuas por se sustentar em dois importantes pressupostos (Buntine, 1994):

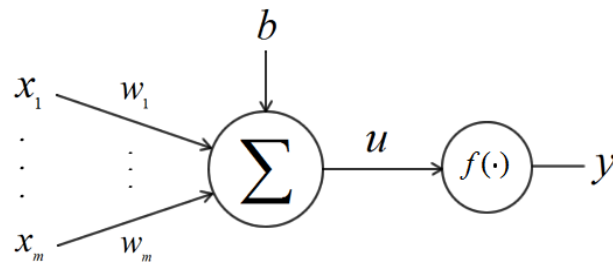
- A suposição que os atributos preditivos são condicionalmente independentes dada a classe.
- Postula-se que nenhum atributo oculto ou subentendido influencia o processo de predição.

#### **2.4.1.3.4 Redes Neurais Artificiais**

Tendo origem na psicologia, as RNA consistem numa técnica que simula o comportamento dos neurônios, ou seja, é uma estrutura computacional interligada que se

comunica entre si, contando com unidades de entrada (input) e saída (output), onde cada ligação apresenta um peso associado, representando a influência que um nó (neurônio) tem com o outro (Souza, 1998).

Na fase de treino, inicia-se o processo de calibração dos pesos associados que vão sendo ajustados de acordo com a saída de informação desejada (Cavalcante, 2014).



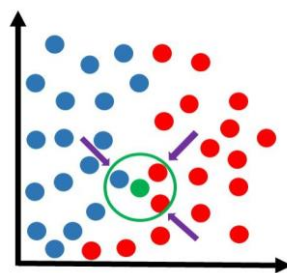
**Figura 7** - Modelo do neurônio artificial

**Fonte:** Haykin, 1998

Segundo Cortez (2002), no neurônio artificial existem três elementos fundamentais, o conjunto de sinapses, o combinador linear e a função de ativação. Onde, as ligações representam as sinapses entre os neurônios, com determinado peso significando a força do sinal enviado, para sinapses de excitação e inibidora são sinais positivos e negativos, respectivamente. O combinador linear ou junção soma, é quando acontece a junção entre os inputs, logo, as entradas são multiplicadas pelo peso exercido na mesma. Em seguida, a função de ativação atua no próprio neurônio processando os valores da combinação linear, transformando-os em uma única saída, que pode ser o resultado do modelo, ou servir de valor de entrada para outro neurônio (Thomé, 2008).

#### 2.4.1.3.5 K-Nearest Neighbours (k-NN)

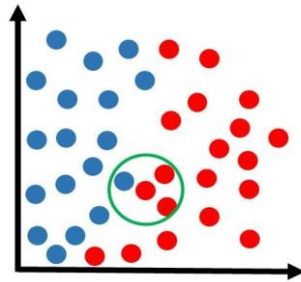
O algoritmo k-NN, também conhecido como, vizinhos mais próximos, consiste na classificação de uma observação desconhecida através da localização em que as k observações estão do item predito.



**Figura 8** - Ilustração do princípio básico da técnica k-NN

**Fonte:** Didática Tech

Neste caso, a figura 8 mostra um exemplo de classificação com duas classes (vermelha e azul), sendo  $k = 3$ , esse valor é definido mediante um conjunto de testes, onde valores altos diminuem a probabilidade de ruídos, em contrapartida, promovem classificações mais complexas. Com isso, através da métrica de distância define-se os vizinhos mais próximos do ponto a ser classificado (ponto verde), classificando o valor desconhecido na mesma classe que surge com maior frequência, representado na figura abaixo (Goldberger et al., 2004).



**Figura 9** - Ilustração da amostra já classificada

**Fonte:** Didática Tech

#### 2.4.1.4 RapidMiner

O RapidMiner Studio versão 10.1.001, consiste num software utilizado para o desenvolvimento de sistemas preditivos, podendo ser aplicado em vários ramos de negócio, a fim de otimizar o processo de tomada de decisão. Esta plataforma ganha espaço pela sua flexibilidade e ambiente gráfico que simplifica as tarefas complexas. Por ser implementado em Java, o RapidMiner torna-se versátil podendo ser operado em qualquer sistema (Bastos, 2015).

De acordo com Bastos (2015), esta ferramenta fornece aos seus utilizadores:

- Uma solução global para o desenho e implementação de um processo de DM.
- Uma interface gráfico muito intuitivo e flexível para o desenho de um processo de DM.
- Mais de 500 operadores que implementam as mais diversas técnicas e algoritmos.
- Acesso as mais diversas fontes de dados: Excel, Access, Oracle, Microsoft SQL Server, SPSS etc.
- Mais de 20 métodos para visualização de dados e modelos.
- Repositórios de processos, dados e metadados.

##### 2.4.1.4.1 Interface do RapidMiner

A interface gráfica disponibiliza várias funcionalidades, permitindo a implementação dos processos sem a necessidade de linguagem de programação. Fornecendo um conjunto

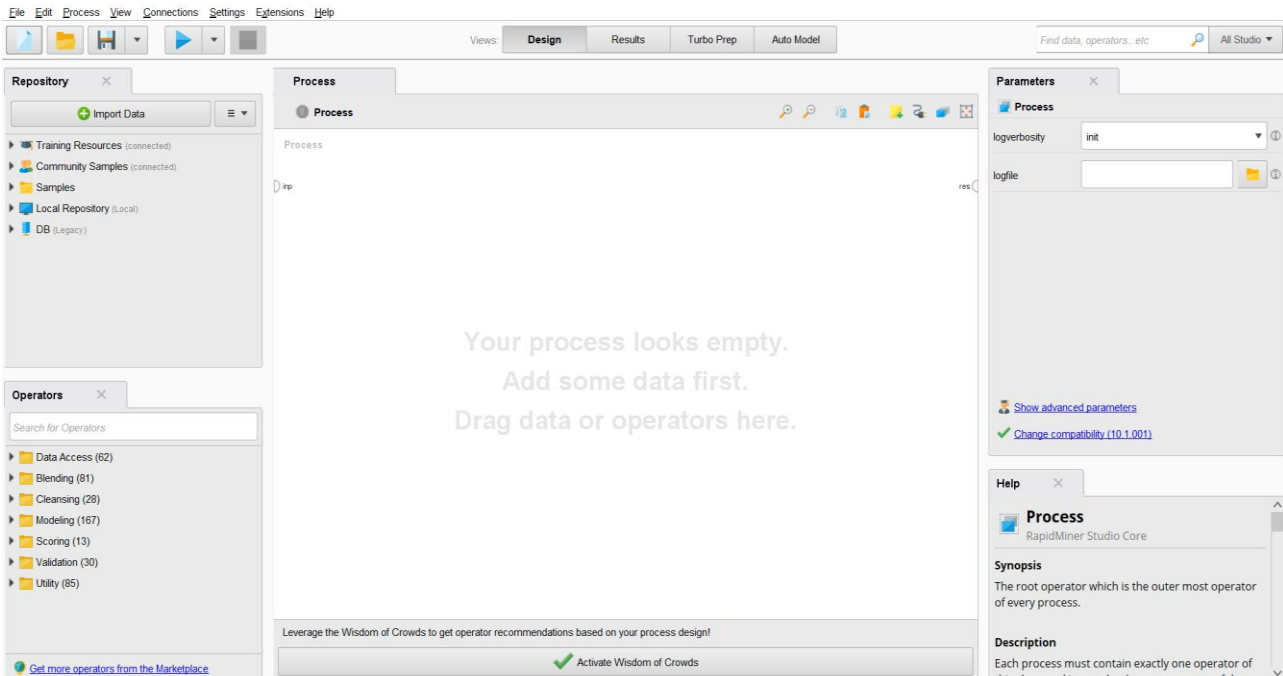
interessante de algoritmos de inteligência artificial que já estão inseridos nos operadores. Sendo oriundo de outros trabalhos de investigação que atuam em problemas específicos (Akthar e Hahne, 2012).

Esta ferramenta permite a definição do tratamento dos dados, com a utilização dos operadores que podem ser individualmente configurados e são incumbidos por:

- Operações de entrada e saída.
- Algoritmos de aprendizagem (supervisionados ou não).
- Funções de seleção e pré-processamento de dados.
- Validação.
- Visualização.

A seguir define-se as três perspectivas da interface:

- **Welcome:** consiste no ecran de boas-vindas.
- **Design:** consiste no ecrã principal onde se constrói todo o processo de DM.
- **Results:** possibilita a análise dos resultados, reunindo todos os outputs do processo de DM, possibilitando a utilização de algumas ferramentas de análise gráfica.



**Figura 10** - Interface do RapidMiner 10.1.001

**Fonte:** Autoria própria, 2023

Segundo Bastos (2015), “Resumidamente, o RapidMiner opera através da elaboração de processos (desenho de projetos de DM). Isto significa que, para realizar uma classificação de dados ou predição de ocorrências nesta ferramenta, é necessário incluir operadores para cada processo que antecede a tarefa final desejada. Alguns operadores possuem parâmetros configuráveis. A lista de operadores, a sua sequência de execução e os parâmetros são armazenados num arquivo XML de fácil visualização e edição.”

### **3 MATERIAL E MÉTODOS**

#### **3.1 Local de Experimentação**

Os dados utilizados neste trabalho são oriundos da Associação Amicus Canis – AMICA, organização sem fins lucrativos, localizada na cidade de Bragança – Portugal. A associação foi fundada em 2011 e desde então os registros dos animais foram feitos em ficheiros físicos, sendo nos últimos anos digitalizados e armazenados em e-mail.

A demanda por cuidado dos gatos é tamanha, e devido à falta de apoio e suporte a associação permanece com dificuldades na organização e controle dos seus registros. Com todo o esforço e dedicação da equipe as informações recolhidas no momento da adoção foram aprimoradas e atualizadas no decorrer dos anos, justificando a falta de informação em alguns registros mais antigos.

##### **3.1.1 Controle Populacional**

Com mais de 80 animais em lares temporários e na instalação da associação, o sistema adotivo continua com problemas, principalmente considerando os animais das colônias que são supervisionadas e controladas na medida do possível, através do processo de CED – Captura, Esterilização e Devolução, ainda assim são muitos animais e continua em expansão.

Tendo em vista as crias de animais errantes, ou seja, animais com tutores que não realizam a castração dos seus animais de estimação, afetando diretamente no controle populacional e aumentando a quantidade de animais errantes que por sua vez, também não terão supervisão das autoridades o que pode acarretar novamente o ciclo de crias indesejadas ou até mesmo, elevar os índices de gatos abandonados.

#### **3.2 Recolha dos Dados**

Portanto, para a realização e criação da base de dados da associação e utilização dos dados, foram utilizadas algumas ferramentas de software, como o *Excel*, o *Access*, assim como o programa *RapidMiner Studio*. Para isso, a forma de envio dos dados foi realizada através do e-mail e fichas impressas.

Com o contacto estabelecido via e-mail, as fichas de adoções disponíveis no e-mail foram-me enviadas, e as mais antigas foram disponibilizadas impressas por parte das

responsáveis da associação. Posteriormente, meu e-mail ficou vinculado e automaticamente recebia as fichas conforme ocorriam novas adoções no decorrer da criação da base de dados.

### 3.2.1 Excel

Para maior facilidade os dados foram transcritos para tabelas *Excel*, nomeadamente uma folha com as informações dos gatos e outra com as informações dos tutores, de acordo com as figuras abaixo.

Nº CHIP	NOME - GATO	SEXO	COLORAÇÃO	CASTRACÃO	DATA NASCIMENTO	LOCAL	DATA ADOÇÃO	IDADE	OBS	ESTADO ADOTIVO
620099200010574	Matilde	Fêmea	Amarelo	Sim	21/11/2020		31/12/2022	2		Sim
620099200021641	Mya	Fêmea	Champagne	Sim	20/02/2023		17/04/2023	2		Sim
620095300158382	Ahsoka Tôna	Fêmea	Branco	Sim	20/04/2022		19/04/2023	12		Sim
620099200021576	Kilua	Macho	Champagne	Sim	20/02/2023		17/04/2023	2		Sim
620099200021579	Mooshie	Macho	Champagne	Sim	20/02/2023		17/04/2023	2		Sim
620095300151287	Alice	Fêmea	Branco/Preto	Sim	01/01/2022		11/04/2023	4		Sim
620095300158500	Scar (Ruby)	Fêmea	Tricolor	Sim	04/08/2021		12/05/2023	10		Sim
620095300158386	Pistache	Macho	Branco/Cinza	Sim	19/03/2023		15/05/2023	2		Sim
620095300158380	Nugget	Macho	Branco/Amarelo	Sim	19/03/2023		15/05/2023	2		Sim
620095300158504	Amora (Gain)	Fêmea	Preto	Sim	19/03/2023		15/05/2023	2		Sim
620095300158391	Lucas	Macho	Branco	Sim	19/03/2023		15/05/2023	2		Sim
620095300158399	Yuki (Hades)	Macho	Champagne	Sim	19/03/2023		15/05/2023	2		Sim
620095300158491	Musathly (Horus)	Macho	Branco/Preto	Sim	19/03/2023		15/05/2023	2		Sim
620099200010786	Francisco (Natan)	Macho	Amarelo	Sim	04/10/2022		13/04/2023	7		Sim
620099200021647	Fifi	Fêmea	Branco/Amarelo	Sim	01/09/2022		20/03/2023	7		Sim
620099200021574	Fofinha	Fêmea	Tricolor	Sim	01/01/2018		20/03/2023	3		Sim
	Roni	Macho	Amarelo	Sim	04/04/2019					Sim
	Chico	Macho	Cinza	Não	16/04/2023					Sim
	Lina	Fêmea	Cinza	Sim	01/09/2021				FeLV +	Não
	Gru	Fêmea	Branco/Cinza	Sim	05/08/2020					Não
	Oreo	Macho	Branco/Preto	Sim	01/08/2019					Não
	Figas	Macho	Amarelo	Sim	01/01/2016	SACOR				Não
	Jonas	Macho	Amarelo	Sim	04/10/2022					Não
	Will	Macho	Tigrado	Sim	25/07/2022					Não
	Willy	Macho	Tigrado	Sim	25/07/2022					Não
	Yan	Macho	Branco/Preto	Sim	25/07/2022					Não
	Renior	Macho	Preto	Sim	01/04/2021					Não
	Misé	Macho	Branco/Amarelo	Sim	15/05/2022	FAT - Dionatan				Não
	Atlas	Macho	Branco/Tigrado	Sim	15/05/2022	SACOR				Não

Figura 11: Planilha com as informações coletadas dos gatos

Fonte: Autoria própria, 2023

Uma vez que um animal dá entrada no sistema, ou seja, é recolhido pela associação por apresentar características procuradas por futuros tutores, os seus dados são registrados, assim como o seu **estado adotivo**, que muda ao ser adotado. Levando automaticamente a atualização da ficha desse animal.

NOME TUTORES	PROFISSÃO	LOCAL DE TRABALHO	MORADA	CÓDIGO POS	LOCALIDAD	CONTACTO	EMAIL	C C	NIF
Lucas Medeiros Hart	Programador	Home Office	Rua Vale D' Álvaro, 62, 6 esq trás	5300-274	Bragança	937240593	lucashart.it@gmail.com	5R907652R	304536296
Paula Grazielly de Souza Hart	Administradora		Rua Vale D' Álvaro, 62, 6 esq trás	5300-274	Bragança	937240592	paulashart@gmail.com	52LL74423	304535907
Carolina Luís Borges Vieira	Estudante		Rua de Montesinho	5300-443	Bragança	934812363	caroluisvieira@gmail.com	152879633ZX3	249577844
Lucia Fernandes	Enfermeira	Hospital (Vila Real)	Rua Dr. Manuel Teixeira, 18	5300-675	Bragança	934748877	luciacosta88@hotmail.com	13560282	213981807
Vera da Conceição Moraes da Costa	Auxiliar lar	Rebardo	Rua dos Retornados, 2, Rebardo	5335-124	Vinhais	934103322	veramorais1980@gmail.com	12691650	223660647
Jozelaine Maria Hipólito da Silva Marques	Limpeza	Prime Clear	Rua Capetão Adriano Peres, 86A	5300-252	Bragança	935982385	hipolitosassessoria@hotmail.com	FV894621	307813320

Figura 12: Planilha com as informações coletadas dos tutores

Fonte: Autoria própria, 2023

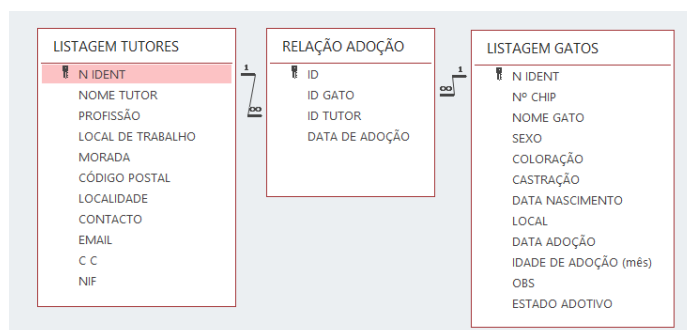
O cadastro do tutor também deve ocorrer para que as informações sejam cruzadas no momento da relação, assim, são recolhidas as informações demonstradas na tabela, a fim de obter maior controle e segurança na adoção.

### 3.3 Access

A escolha deste *software* partiu da praticidade do uso e disponibilidade por parte da associação. Assim como, a ampla possibilidade de opções e ferramentas para garantir a qualidade das interfaces, sendo organizados e dinâmicos.

#### 3.3.1 Modelo Relacional

Uma base de dados é composta por um conjunto de tabelas e associações entre as tabelas. A associação entre os dados é o ponto forte dos sistemas relacionais. As tabelas são formadas por linhas e colunas onde figuram os dados. Numa base de dados relacional os dados estão todos representados como valores nas colunas das tabelas (Caldeira, 2004).



**Figura 13** - Esquema relacional da BD funcional do sistema

**Fonte:** Autoria própria, 2023

Como já mencionado, o registo no sistema deve ocorrer para os animais e adotantes quando necessário, desta forma através da relação entre tabelas, o sistema junta as informações criando o arquivo dos animais adotados com todas as suas informações, bem como, as informações do seu respectivo tutor(a).

#### 3.3.2 Tabelas

Foram criadas no total três tabelas como suporte para a criação da base de dados, nomeadamente, a tabela de registros dos gatos, a dos registros dos tutores e a tabela de relação entre as duas, a partir delas que se consegue dar origem aos formulários e garantir o bom funcionamento do sistema. Abaixo segue a estrutura de cada tabela, ou seja, com quais informações foram criadas (nome do campo), bem como, a configuração dos dados (tipo de dados). Ressaltando a importância dessas configurações para o bom funcionamento da base de dados.

**Tabela 1:** Estrutura da tabela de registro dos gatos

Nome do campo	Tipo de dados
N IDENT	Numeração Automática
Nº CHIP	Número
NOME GATO	Texto Curto
SEXO	Texto Curto
COLORAÇÃO	Texto Curto
CASTRACÃO	Texto Curto
DATA NASCIMENTO	Data/Hora
LOCAL	Texto Curto
DATA ADOÇÃO	Data/Hora
IDADE DE ADOÇÃO (mês)	Número
OBS	Texto Curto
ESTADO ADOTIVO	Texto Curto

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Número Chip (Nº CHIP); Observação (OBS)

As especificações foram ajustadas de acordo com a maior facilidade de leitura e utilização dos dados, as informações coletadas e selecionadas foram de acordo com as disponibilizadas pela associação, de modo que fosse favorável para atender os objetivos do projeto.

**Tabela 2:** Estrutura da tabela de registro dos tutores

Nome do campo	Tipo de dados
N IDENT	Número
NOME TUTOR	Texto Curto
PROFISSÃO	Texto Curto
LOCAL DE TRABALHO	Texto Curto
MORADA	Texto Curto
CÓDIGO POSTAL	Texto Curto
LOCALIDADE	Texto Curto
CONTACTO	Texto Curto
EMAIL	Texto Curto
C C	Texto Curto
NIF	Número Grande

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Cartão de Cidadão (CC); Número de Identificação Fiscal (NIF)

Considerando os mesmos requisitos para construir a tabela apresentada, com as informações disponibilizadas pela associação.

**Tabela 3:** Estrutura da tabela de relação

Nome do campo	Tipo de dados
ID	Número
ID GATO	Número
ID TUTOR	Número
DATA DE ADOÇÃO	Data/Hora

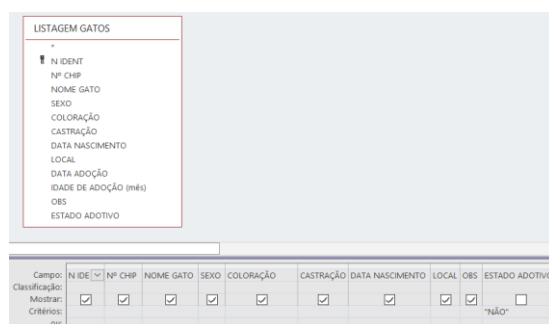
**Fonte:** Autoria própria, 2023

Identificação (ID)

Esta, foi criada para automatizar e garantir a funcionalidade do sistema, facilitando a separação de animais adotados e disponíveis. Contém a chave primária das tabelas 1 e 2, promovendo a ligação e relação entre ambas.

### 3.3.3 Consultas

As consultas são criadas com o propósito de filtrar as informações das tabelas, ou seja, ver apenas o que lhe interessa de forma rápida e prática. Assim como, permitir a criação dos relatórios. Essa filtragem consiste em solicitações efetuadas à base de dados, que foram escolhidas de acordo com o maior interesse e demanda da associação, a fim de facilitar o controle dos animais.



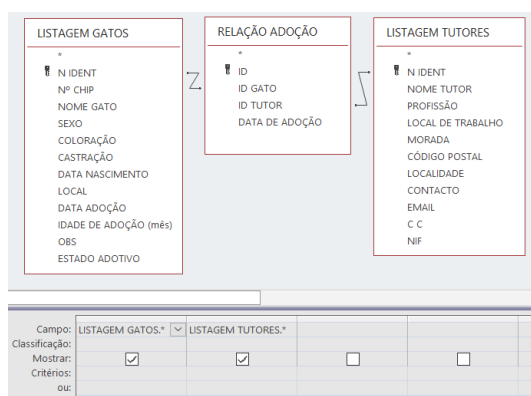
**Figura 14:** Consulta dos gatos para adoção em modo de estrutura

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Número Chip (N° CHIP); Observação (OBS)

Para a criação da consulta apresentada, selecionou-se a tabela desejada e filtrou-se os campos que eram interessantes aparecer na consulta, além de criar critérios para auxiliar na criação da consulta.

Neste caso, foi utilizado a expressão com aspas, que seleciona todos os registros que possuem a respectiva palavra inserida na expressão, ou seja, no campo estado adotivo, foi filtrado os registros que continham a palavra “não”, logo, animais não adotados. Criando uma consulta específica para esses animais.

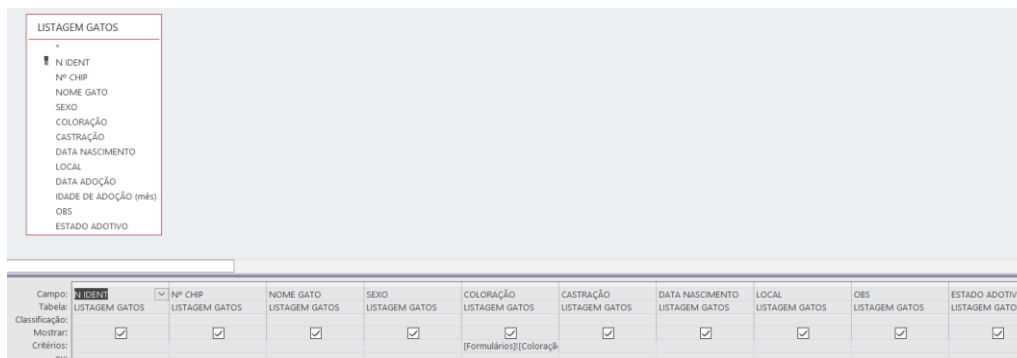


**Figura 15:** Consulta dos gatos adotados em modo de estrutura

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Número Chip (N° CHIP); Observação (OBS); Identificação (ID); Cartão de Cidadão (CC); Número de Identificação Fiscal (NIF)

Para essa consulta, foi necessário a utilização da tabela dos gatos, e da tabela dos tutores, possibilitando a junção das informações de ambas as tabelas. E para automatizar as informações presentes nesta consulta, usou-se a tabela de relação entre a listagem dos gatos e tutores.

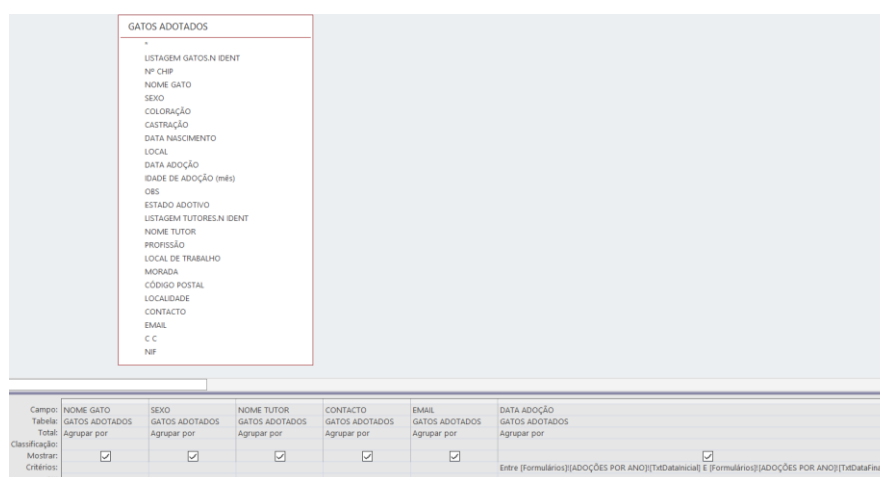


**Figura 16:** Consulta por coloração em modo de estrutura

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Número Chip (Nº CHIP); Observação (OBS)

Objetivando a facilidade no controle dos animais, essa consulta foi criada, a fim de facilitar a busca por animais de uma determinada cor, considerando que tal característica é uma das principais razões para adoção. Para isso, foram utilizadas as funções de parâmetros na característica **coloração**, e para melhor visualização e aparência, foram criados um formulário e um relatório com base nesta consulta. Detalhados a seguir nos itens 3.3.4 e 3.3.5 respectivamente.



**Figura 17:** Consulta de adoções por ano em modo de estrutura

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Número Chip (Nº CHIP); Observação (OBS); Cartão de Cidadão (CC); Número de Identificação Fiscal (NIF)

Criada para o controle das taxas de adoções por ano, usufruindo da função **entre com parâmetro**, para data inicial e data final, aplicada na característica **data adoção**, da mesma maneira foram criados um formulário e relatório para facilitar a visualização e dinâmica das apresentações das informações.

### 3.3.4 Formulários

Como forma de dinamizar as informações, os formulários são criados, bem como, para facilitar a visualização e organização informativa. Contribui com interfaces personalizados e criativos, tornando o projeto único com os seus diferenciais. Além de permitir a realização de algumas funções através dos botões criados, permitindo passar de uma ficha para outra, pesquisar pelo nome ou qualquer outra característica uma ficha específica, adicionar novos registros, excluir registros e imprimir as fichas individuais, se necessário.

AMICA AMICUS CANIS		FICHA INDIVIDUAL - PARA ADOÇÃO		✕	
N IDENT			80		
Nº CHIP					
NOME GATO			Xavier		
SEXO			Macho		
COLORAÇÃO			Branco/Cinza		
CASTRACÃO			Sim		
DATA NASCIMENTO			08/03/2022		
LOCAL			SACOR		
OBS					
ESTADO ADOTIVO			Não		

◀ ▶ PESQUISAR ADICIONAR EXCLUIR IMPRIMIR

**Figura 18:** Formulário dos gatos para adoção – Gato Xavier

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Número Chip (Nº CHIP); Observação (OBS)

Todos os formulários desta categoria contam com as devidas informações presentes na figura acima, logo, quando um animal da entrada na associação é para esta categoria que o animal entra. O formulário dispõe de botões que o tornam prático e facilitam o uso e funcionamento no arquivo.



FICHA INDIVIDUAL - TUTORES

N IDENT	80
NOME TUTOR	Natasha Gabrielly Porrua
PROFISSÃO	Mestranda
LOCAL DE TRABALHO	IPB
MORADA	Av. Dr. Franc. Sá Carneiro, 103 ap -2 centro
CÓDIGO POSTAL	5300-252
LOCALIDADE	Bragança
CONTACTO	912772978
EMAIL	porruanatasha@gmail.com
C C	9P6K75087
NIF	309756464

◀ ▶
PESQUISAR
ADICIONAR
EXCLUIR
IMPRIMIR

**Figura 19:** Formulário dos tutores – Tutora Natasha

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Cartão de Cidadão (CC); Número de Identificação Fiscal (NIF)

O formulário dos tutores conta com suas principais informações, facilitando a organização do sistema adotivo e tornando mais eficiente a resolução de casos de desaparecimento dos gatos.



FICHA INDIVIDUAL - ADOTADOS

N IDENT	80	N IDENT	80
Nº CHIP	620095300103541	NOME TUTOR	Natasha Gabrielly Porrua
NOME GATO	Xavier	PROFISSÃO	Mestranda
SEXO	Macho	LOCAL DE TRABALHO	IPB
COLORAÇÃO	Branco/Cinza	MORADA	Av. Dr. Francisco Sá Carneiro, apto -2 centro
CASTRACÃO	Sim	CÓDIGO POSTAL	5300-252
DATA NASCIMENTO	08/03/2022	LOCALIDADE	Bragança
DATA ADOÇÃO	17/09/2022	CONTACTO	912772978
IDADE DE ADOÇÃO (mês)	7	EMAIL	porruanatasha@gmail.com
OBS		C C	9P6K75087
		NIF	309756464


◀ ▶
PESQUISAR
ADICIONAR
EXCLUIR
IMPRIMIR

**Figura 20:** Formulário dos gatos adotados – Gato Xavier

**Fonte:** Autoria própria, 2023

Número de Identificação (N IDENT); Número Chip (Nº CHIP); Observação (OBS); Cartão de Cidadão (CC); Número de Identificação Fiscal (NIF)

Este formulário contém a combinação das informações dos gatos e dos seus tutores conforme exemplificado nas tabelas acima, portanto, quando a estado adotivo de um gato passa de **não adotado** para **adotado**, esse animal passa automaticamente para esta categoria, onde, juntamente das informações do seu tutor (a), este formulário é criado.

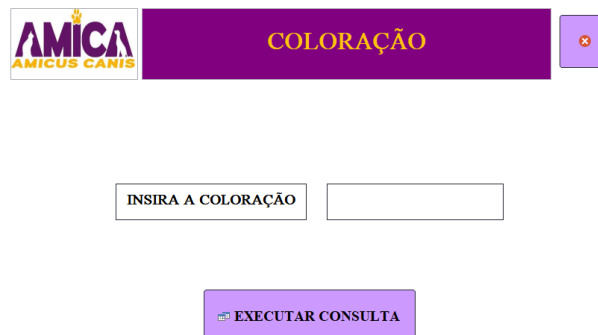


AMICA AMICUS CANIS	ADOÇÕES POR ANO	✕
DATA INICIAL	<input type="text"/>	
DATA FINAL	<input type="text"/>	
<input type="button" value="EXECUTAR CONSULTA"/>		

**Figura 21:** Formulário para consulta de adoções por ano

**Fonte:** Autoria própria, 2023

Esta função foi criada para ter acesso e maior controle perante as adoções realizadas dentro de um determinado período. Com isso, foi utilizado os dados reunidos na consulta apresentada na figura 17, para criar um relatório (tabela 4) com essas informações.



AMICA AMICUS CANIS	COLORAÇÃO	✕
INSIRA A COLORAÇÃO	<input type="text"/>	
<input type="button" value="EXECUTAR CONSULTA"/>		

**Figura 22:** Formulário para consulta de animais com a mesma coloração

**Fonte:** Autoria própria, 2023

Este formulário foi criado para facilitar a visualização e controle dos animais com a mesma coloração de pelagem, visto que é uma característica importante e relevante na hora da adoção. Para isso, foi criado esta função que como a anterior, utiliza os dados reunidos na consulta relatada na figura 16, para criar um relatório com tais informações (tabela 5). Ou seja, são relatórios executados através de um formulário, com as informações das respectivas consultas.



**Figura 23:** Formulário criado como página de arranque

**Fonte:** Autoria própria, 2023

O formulário de arranque foi criado com a intenção de facilitar a utilização do sistema gestor da base de dados, para isso, foram criados botões com funções específicas que dão acesso a todo o arquivo. Além de tornar a interface da BD mais dinâmica e autêntica, permite a navegação e atualização da BD de forma clara e prática.

### 3.3.5 Relatórios

Os relatórios servem para gerar documentos organizados e específicos das informações mais relevantes de uma BD, disponibilizando de marcadores de data e horário, o que auxilia na organização dos arquivos, além de permitir a possibilidade de imprimir tais arquivos caso seja necessário.

**Tabela 4:** Relatório das adoções nos meses de maio e junho

AMICA AMICUS CANIS		GATOS ADOTADOS - POR ANO				28 de agosto de 2023		
						18:57:01		
Nº CHIP	NOME GATO	SEXO	NOME TUTOR	CONTACTO	EMAIL	DATA ADOÇÃO		
620095300158504	Amora	Fêmea	Carolina Luis Borges Vieira	934812363	caroluisvieira@gmail.com	15/05/2023		
620095300158491	Musathy	Macho	Jozelaine Maria Hipólito Marques	935982385	hipolitosassessoria@hotmail.com	15/05/2023		
620095300158399	Yuki	Macho	Vera da Conceição Morais da Costa	934103322	veramorais1980@gmail.com	15/05/2023		
620095300158391	Lucas	Macho	Lucia Fernandes	934748877	luciacosta88@hotmail.com	15/05/2023		
620095300158386	Pistache	Macho	Lucas Medeiros Hart	937240593	lucashart.it@gmail.com	15/05/2023		
620095300158380	Nugget	Macho	Paula Grazielly de Souza Hart	937240592	paulashart@gmail.com	15/05/2023		
620095300158228	Pipoca	Fêmea	Maria Albertina Santos Rodrigues	968511722	engmaria.rodrigues@gmail.com	15/05/2023		
620095300167105	Chico	Macho	Hellena da Silva Campos	924466836	camposhellena@outlook.com	25/06/2023		
620095300167065	Rony	Macho	Natasha Gabrielly Porrua	912772978	porruanatasha@gmail.com	25/06/2023		

TOTAL 9

IMPRIMIR

Fonte: Autoria própria, 2023

Este relatório é o resultado da execução do formulário apresentado na figura 21, facilitando a visualização e simplificando as buscas. Qualquer período pode ser consultado, melhorando o controle das adoções, assim como, promover conhecimento dos períodos de maior adoção, permitindo a criação de estratégias adotivas. Aproveitando o formato do documento, em caso de necessidade o relatório pode ser impresso.

**Tabela 5:** Relatório dos animais com coloração amarela

AMICA AMICUS CANIS		RELATÓRIO POR COLORAÇÃO				28 de agosto de 2023		
						19:05:38		
Nº CHIP	NOME	SEXO	COLORAÇÃO	CASTRACÃO	D. NASCIMENTO	LOCAL	OBS	ESTADO ADOT.
	Fisgas	Macho	Amarelo	Sim	01/01/2016	SACOR		Não
	Gaspar	Macho	Amarelo	Sim	05/05/2017			Não
620095300167642	Chico	Macho	Amarelo	Sim	21/09/2017			Sim
991001002873793	Kiko	Macho	Amarelo	Sim	12/02/2019			Sim
620095300167065	Rony	Macho	Amarelo	Sim	04/04/2019			Sim
991001002040104	Cookie	Macho	Amarelo	Não	25/06/2019			Sim
991001003627197	Neo	Macho	Amarelo	Sim	21/07/2020			Sim
620099200010574	Matilde	Fêmea	Amarelo	Sim	21/11/2020			Sim
	Derek	Macho	Amarelo	Sim	04/10/2022			Não
	Jonas	Macho	Amarelo	Sim	04/10/2022			Não
620099200010786	Francisco	Macho	Amarelo	Sim	04/10/2022			Sim

TOTAL 11

IMPRIMIR

Fonte: Autoria própria, 2023

Tendo em vista que a coloração é uma característica que os futuros adotantes levam em consideração, foi pertinente a criação de um relatório específico para tal característica. Este, que é o resultado da execução do formulário demonstrado na figura 22.

**Tabela 6:** Relatório dos animais para adoção

AMICA AMICUS CANIS		GATOS PARA ADOÇÃO					28 de agosto de 2023		20:11:23	
N IDENT	Nº CHIP	NOME GATO	SEXO	COLORAÇÃO	CASTRAÇÃO	D. NASCIMENTO	LOCAL	OBS		
117		Lina	Fêmea	Cinza	Sim	01/09/2021		FeLV +		
118		Gru	Fêmea	Branco/Cinza	Sim	05/08/2020				
119		Oreo	Macho	Branco/Preto	Sim	01/08/2019				
120		Fisgas	Macho	Amarelo	Sim	01/01/2016	SACOR			
121		Jonas	Macho	Amarelo	Sim	04/10/2022				
122		Will	Macho	Tigrado		25/07/2022				
123		Willy	Macho	Tigrado		25/07/2022				
124		Yan	Macho	Branco/Preto		25/07/2022				
125		Renior	Macho	Preto	Sim	01/04/2021				
126		Misé	Macho	Amarelo/Branco	Sim	15/05/2022	FAT			
127		Atlas	Macho	Branco/Tigrado	Sim	15/05/2022	SACOR			
128		Páscoa	Fêmea	Cinza	Sim	01/01/2020	SACOR			
129		Fred	Macho	Preto	Sim	01/06/2021				
130		Milton	Macho	Branco/Preto	Sim	01/05/2022	SACOR			
131		Francisco	Macho	Tigrado	Sim	05/04/2021				
132		Pimenta	Fêmea	Preto		19/08/2022				
133		Canela	Fêmea	Tricolor	Sim	19/08/2022	SACOR	Perna direita traseira amputada		
134		Latoya	Fêmea	Branco/Preto	Sim	10/04/2019				
135		Rômulo	Macho	Tigrado	Sim	24/07/2021				
136		Rebbie	Fêmea	Branco/Preto	Sim	10/04/2019				

Fonte: Autoria própria, 2023

Documento feito para facilitar a organização das fichas dos animais disponíveis para adoção, otimizando a BD.

**Tabela 7:** Relatório dos animais adotados

AMICA AMICUS CANIS		GATOS ADOTADOS										30 de agosto de 2023		18:20:06	
Nº CHIP	NOME GATO	SEXO	COR	CASTRAC.	DATA NASCIMENTO	DATA ADOÇÃO	IDADE (mês)	NOME TUTOR	PROFISSÃO	LOCAL DE TRABALHO	MORADA	LOCALIDADE	CONTACTO		
620098500147694	Luna	Fêmea	Cinza Tigrado	Sim	02/07/2020	23/12/2020	6	Ana Isabel Bento Carvalho	Taxista	Taxista - Bragança	Rua São João Bosco, 1, r/c esq	Bragança	938491052		
991001003627151	Poppy	Fêmea	Branco/Bege	Sim	01/10/2020	24/12/2020	3	Bruna Alexandra Campos	Médica Dentista	Clínica Dentária Lucas Taden	Rua Amália Rodrigues, 10, hab 72	Bragança	916434826		
963020010561586	Lili	Fêmea	Branco/Tigrado	Não	12/10/2020	12/01/2021	3	Carla Marina Esteves	Professora	ES/3 Abade Baçal	Av. Dinastia de Bragança, 3, 3º Dto	Bragança	917822939		
953000010561476	Lili	Fêmea	Branco/Tigrado	Sim	10/08/2020	19/01/2021	5	Antónia Conceição Ferreira	Administrativo	Balarista Sagmas (Fidelidade)	Av. Das Forças Armadas Nivel	Bragança	933167847		
620098500173319	Preta	Fêmea	Preto	Sim	08/09/2020	08/02/2021	5	Vicente Santos Fernandes	Advogado	Escritório	Rua Luis Lobo, 2	Bragança	966516720		
620098500139413	Gigi	Fêmea	Branco/Bege	Sim	01/10/2020	19/03/2021	5	Nelson Filipe Cruz	Assistente Operacional	Hospital Bragança	Rua Camilo Castelo Branco, 98	Bragança	910264002		
620098500139404	Bartolomeu	Macho	Amarelo/Branco	Sim	05/10/2020	19/03/2021	5	Nelson Filipe Cruz	Assistente Operacional	Hospital Bragança	Rua Camilo Castelo Branco, 98	Bragança	910264002		
991001003627197	Neo	Macho	Amarelo	Sim	21/07/2020	21/03/2021	8	Ana Isabel Bento Carvalho	Taxista	Taxi - Bragança	Rua São João Bosco, 1, r/c esq	Bragança	987562415		
620095300075895	Nevinha	Macho	Branco	Sim	01/01/2020	25/03/2021	14	Sara Vicente Leonor	GNR	Bragança	Rua do Souto, 25, Sanil. BFC	Sanil	963101807		
620095300075886	Pipoca	Fêmea	Branco/Preto	Sim	15/08/2020	25/03/2021	7	Teresa Augusta Rodrigues Neves	Museóloga	Câmara Municipal de Vinhais	Av. Padre Firmino Martins, 16, ap 3º esq	Vinhais	938037926		
620098500224568	Kilo	Macho	Branco/Cinza	Sim	01/04/2021	11/06/2021	2	Susana Cristina Vaz Salgado	Desempregada	-	Eng. Oliveira Bras. 1, 4º esq	Bragança	930407158		
620098500224567	João	Macho	Preto	Sim	05/04/2021	11/06/2021	2	Soraia de Lurdes F. Raimundo	Investigadora	IPB	Rua da Paz, 170, Mirandela (Carvelhais)	Mirandela	914789934		

Fonte: Autoria própria, 2023

Relatório dos animais que já foram adotados, com informações dos seus respectivos tutores, facilitando a contabilização para o sistema adotivo.

### 3.4 RapidMiner Studio

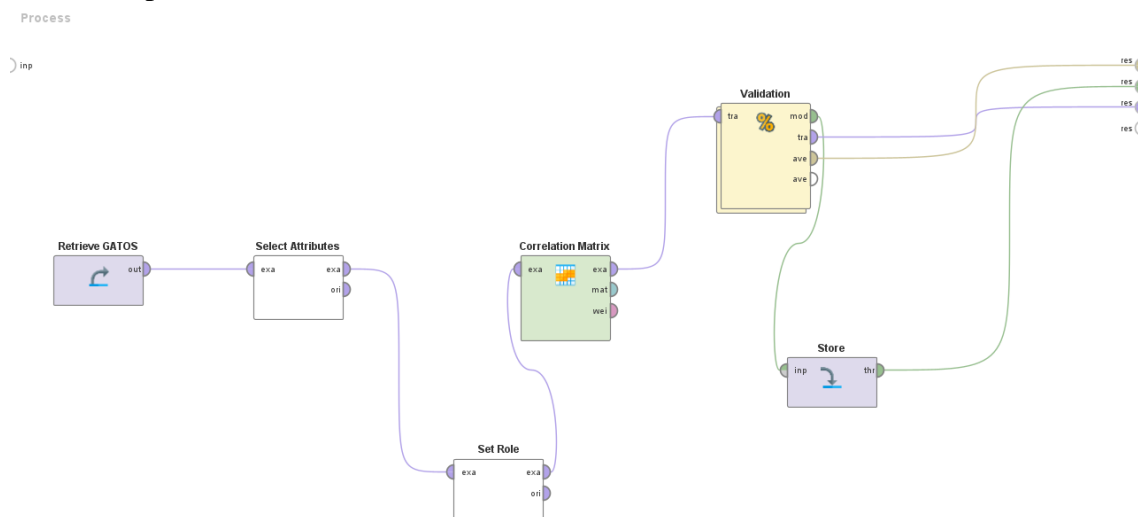
Considerando a praticidade e capacidade de criar modelos preditivos em diferentes áreas de negócios, este *software* foi escolhido para o desenvolvimento do presente trabalho, com a sua flexibilidade e ambiente gráfico possibilitou a realização de *interfaces* intuitivos e simplificados.

#### 3.4.1 Modelo de Predição

Consiste em uma função matemática que pode ser aplicada em uma BD, a fim de identificar padrões capazes de apontar tendências, neste caso, de forma estatística e matemática promove a previsão do estado adotivo dos gatos, ou seja, identifica a probabilidade deste animal ser ou não adotado.

##### 3.4.1.1 Criação do Modelo Preditivo

Implementado no *software RapidMiner*, o modelo é criado e validado tendo uma variável alvo, sendo ela, a probabilidade de adoção dos gatos. O processo concentra-se em otimizar as ações, intencionando um padrão de comportamento futuro com base em casos ocorridos no passado (Bastos, 2015).



**Figura 24:** Processo de criação do modelo preditivo

**Fonte:** Autoria própria, 2023

A definição e explicação do processo é realizada segundo Bastos (2015):

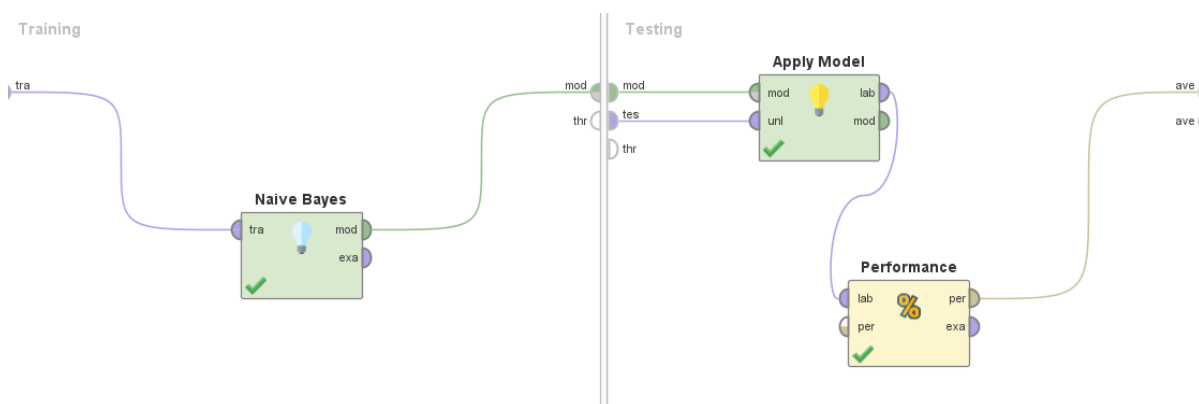
O processo inicia-se com a inserção da base de dados, neste caso foi exportado dados da BD e criado um documento Excel, unificando os dados dos animais adotados e não adotados, onde o campo “Retrieve GATOS” importou os dados presentes no documento Excel. Na sequência, utilizando o recurso “Select Attributes” realizou-se a filtragem dos atributos, ou seja, selecionou-se somente os que são relevantes para a criação do modelo.

Na fase seguinte, com o operador “Set Role” foi definido o atributo alvo do estudo, ou seja, o atributo Adotado. Na sequência, é aplicado o operador “Correlation Matrix” responsável por dispor de uma técnica estatística que demonstra a intensidade da relação entre dois atributos, indicando a correlação entre os atributos, auxiliando e refinando a criação do modelo.

Através do operador “Validation” foi possível efetuar uma aprendizagem supervisionada para a geração do modelo preditivo, que foi facilitada pelos pesos oriundos da matriz de correlação. Nesta fase, os dados foram embaralhados a fim de diminuir as taxas de erro, além de serem divididos automaticamente em 70% para treino e 30% para teste, ou seja, respectivamente para aprendizagem e validação do modelo. A fim de facilitar a utilização do modelo em futuras previsões, este modelo é gravado em um repositório, através do operador “Store”.

#### 3.4.1.1.1 Validação e Aplicação do Modelo Preditivo

Abaixo segue a configuração do operador “Validation”, bem como, a representação do algoritmo utilizado na geração do modelo.



**Figura 25:** Configuração e aplicação do operador validation

**Fonte:** Autoria própria, 2023

A justificativa e vantagens consideradas para a utilização do algoritmo de DM, “*Naive Bayes*”, foi devido ao fato de ser uma técnica que para a obtenção de resultados satisfatórios, não necessita de grandes volumes de dados, além de ser possível que a seleção dos campos utilizados para a geração do modelo preditivo não ter de ser obrigatoriamente a mesma para a geração da predição.

Na fase de treino como já mencionado, somente 70% dos dados são utilizados, posteriormente com a utilização do operador “*Apply Model*”, o modelo é definitivamente gerado e aplicado à percentagem de dados para teste (30%), medindo o seu desempenho, possibilitando a verificação do grau de certeza das previsões efetuadas.

accuracy: 84.78%

	true sim	true Não	class precision
pred. sim	31	5	86.11%
pred. Não	2	8	80.00%
class recall	93.94%	61.54%	

**Figura 26:** Matriz de Confusão do Modelo

**Fonte:** Autoria própria, 2023

Uma matriz de confusão consiste numa tabela que indica os erros e acertos do modelo gerado, comparando com o resultado esperado (ou etiquetas/*labels*). Nas linhas podemos analisar os valores reais e nas colunas os valores detectados. Cruzando a linha “*pred. sim*” com a coluna “*true sim*” temos os verdadeiros positivos (31) e cruzando a linha “*pred. não*” temos os verdadeiros negativos (8). Os valores falsos positivos (2) surgem do cruzamento da linha “*pred. não*” com a coluna “*true sim*” que corresponde ao erro em que o modelo previu a classe Positivo quando o valor real era classe Negativo. Os valores falsos negativos (5) surgem do cruzamento da linha “*pred. sim*” com a coluna “*true não*” que corresponde ao erro em que o modelo previu a classe Negativo quando o valor real era classe Positivo.

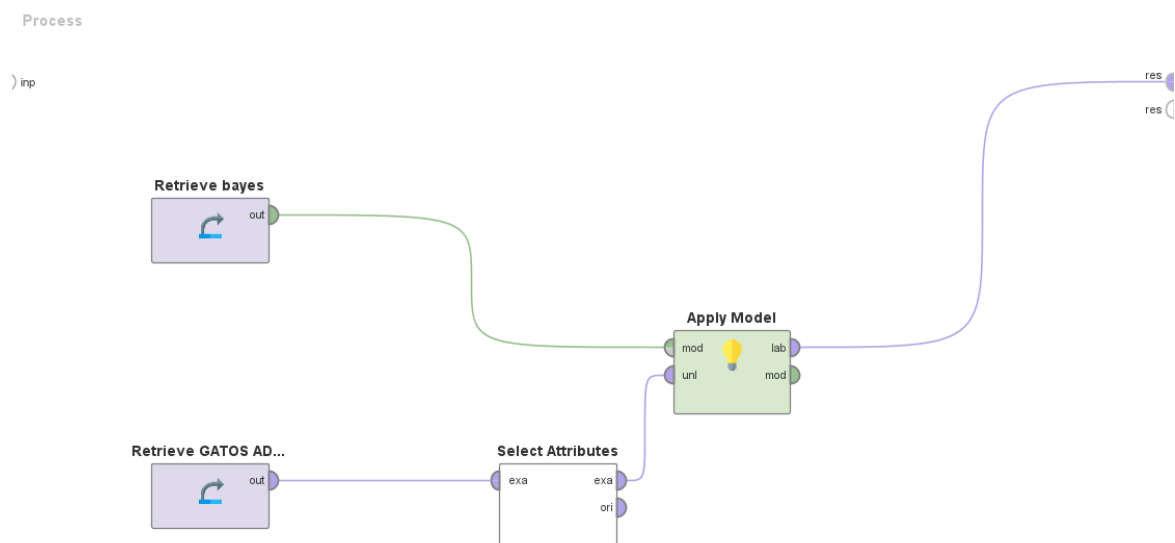
Ao ser feita a contagem de todos esses termos e obter a matriz de confusão, é possível calcular métricas de avaliação para a classificação:

- **Acurácia:** indica uma performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente (84,78%);
- **Precisão:** dentre todas as classificações de classe Positivo e Negativo que o modelo fez, quantas estão corretas (86,11% e 80,00% respetivamente);

- **Recall/Revocação/Sensibilidade:** dentre todas as situações de classe Positivo e Negativo como valor esperado, quantas estão corretas (93,94% e 61,54% respectivamente).

### 3.4.1.2 Caso de Estudo

Considerando o tamanho da BD e o objetivo final do presente trabalho, foi criado um caso de estudo simples, a fim de verificar a eficiência do modelo preditivo. Para isso, criou-se três animais fictícios com as suas determinadas informações, e aplicou-se o modelo preditivo, como demonstrado na figura abaixo.



**Figura 27:** Teste realizado com o modelo

**Fonte:** Autoria própria, 2023

Esta figura demonstra o teste realizado a fim de conferir a eficiência do modelo. Logo, como já mencionado, criou-se três animais fictícios presentes no “*Retrieve GATOS ADOTAR*”, onde não obtinham resultados quanto ao estado adotivo, foram associados ao algoritmo presente no operador “*Retrieve bayes*”, tendo os seus campos filtrados e selecionados através do operador “*Select Attributes*”, para então ser aplicado o modelo preditivo via o operador “*Apply Model*”.

### 3.4.1.2.1 Resultados Obtidos

Tornando o modelo válido e eficaz, obteve-se resultado em relação ao caso prático, certificando a funcionalidade do projeto.

**Tabela 8:** Resultados do caso de estudo

Row No.	prediction(Adotado)	confidence(sim)	confidence(Não)	Nº CHIP	NOME GATO	SEXO	COLORAÇÃO	CASTRACÃO	DATA NASCIMENTO	LOCAL
1	sim	0.999	0.001	?	Tareco	Macho	Preto	?	Jul 4, 2017	?
2	sim	1.000	0.000	?	Pintas	Macho	Preto/Branco	?	Nov 12, 2021	?
3	sim	1.000	0.000	?	Nica	Fêmea	Tricolor	?	Sep 4, 2022	?

**Fonte:** Autoria própria, 2023

Os resultados foram positivos quanto a possibilidade de adoção dos gatos fictícios, resultando em 100% de certeza para adoção desses animais.

## **4 RESULTADOS E DISCUSSÃO**

### **4.1 Sistema de Gestão de Base de Dados**

Através da aplicação *Access*, foi criado do zero uma base de dados para a Associação Amicus Canis, a BD conta com poucos registros, justamente pelo fato de que os registros são feitos com documentos impressos, podendo ocorrer a perda de alguns ou até mesmo de alguma forma ser ou se tornar ilegível. De toda maneira, a base de dados foi criada sendo válida e funcional, permanecendo como ferramenta para a associação, melhorando e tornando mais eficiente o controle dos registros dos animais.

Com o modelo relacional conectado e operando, a base de dados tornou-se funcional, sendo possível a realização de processos através da manipulação e utilização dos dados que estarão registrados e guardados de forma segura e organizada, podendo ser utilizada para otimizar o sistema adotivo da associação, facilitando a tomada de decisão em relação as adoções.

### **4.2 Aplicação do Data Mining**

Considerando o objetivo de prever futuras adoções para a associação, utilizou a tarefa de previsão associada a uma aprendizagem supervisionada, onde ocorre a divisão dos dados, realizando a fase de treino e de validação, neste caso com divisão 70% e 30% respectivamente (Cortez, 2002). Desta forma, atingiu-se uma performance de 84,78% de precisão quanto aos resultados previstos, para isso, foi manuseado o algoritmo *Redes Bayes*, que consiste na detecção numérica da probabilidade que um evento pode ocorrer, baseada na probabilidade de um evento que já ocorreu (Camilo, 2009).

O sistema de predição criado com a performance atingida e apresentada, corresponde de forma funcional e válida, uma vez que tenha passado por todas as etapas necessárias de criação, respeitando um conjunto de processos cruciais e fundamentais para a validação do modelo preditivo.

### **4.3 Resultados Obtidos**

A fim de justificar o tópico 3.4.1.2.1, com resultados de 100% de adoção para os animais criados como estudo de caso, ressalta a quantidade de dados na BD, ou seja, a baixa quantidade de dados de animais adotados e não adotados interfere na veracidade dos valores,

uma vez que não se obtém 100% de certeza em relação a nada. Entretanto, o modelo de predição apresenta-se como funcional e válido, significando que com o passar do tempo e aumento da base de dados os o modelo preditivo irá refinar, tornando os resultados mais precisos e mais confiáveis.

## **5 CONCLUSÕES**

A criação do sistema gestor de base de dados foi concluída de forma eficiente, justamente pelos dados ficarem seguros e organizados com facilidade de acesso, além da aplicação do data mining, que promoveu a possibilidade da criação do modelo preditivo capaz de prever a possibilidade de adoção dos animais resgatados pela associação. Neste caso, os alcances foram de encontro com os objetivos pré-estabelecidos no planejamento do presente projeto, validando e justificando a utilização do trabalho realizado.

### **5.1 Considerações Finais**

A criação desse sistema automatizou e facilitou o armazenamento dos dados, assim como, aumentou a praticidade de acesso as fichas de adoções. Além de possibilitar a previsão quanto as chances de adoção dos animais, pois o sistema apresenta-se como funcional, devemos considerar a veracidade dos resultados obtidos devido à quantidade de dados atuais da base de dados. Entretanto de todo modo, os dados podem ser utilizados, gerando resultados que serão ainda mais confiáveis conforme o crescimento da base de dados.

Por fim, com a criação de um sistema autêntico e dinâmico facilitou a visualização das regras e padrões traçados como características importantes e consideradas no momento da adoção, auxiliando no processo de tomada de decisão, bem como, promovendo facilidades nas estratégias estabelecidas para o sistema adotivo.

### **5.2 Recomendação para Trabalho Futuro**

Considerando a não automatização no complexo de exportação dos dados para o arquivo de suporte ao sistema de predição, vale ressaltar a importância desse mecanismo para a facilidade e praticidade na manipulação do sistema. Logo, recomenda-se a criação de automatismo de ligação entre a base de dados e o sistema de predição, que além de fornecer praticidade, colabora com a diminuição na taxa de erros na exportação.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

- Akthar, F. & Hahne, C. (2012). *RapidMiner 5 Operator Reference*.
- Amo, S. (2004). *Técnicas de mineração de dados*.
- Amorim, T. (2006). *Conceitos, técnicas, ferramentas e aplicações de Mineração de Dados para gerar conhecimento a partir de bases de dados*. Monografia, Bacharel em Ciência da Computação, Universidade Federal de Pernambuco.
- Apte, C., et al. (2002). *Business Applications of Data Mining*. ACM, 45(8): p. 49-53.
- Bastos, P. M. L. (2015). *Sistema de predição de avarias em máquinas de unidades fabris globalmente dispersas*. Universidade do Minho, Portugal, Tese de Doutorado.
- Boente, A. N., Goldschmidt, R. R., & Estrela, V. V. (2008). *Uma metodologia de suporte ao processo de descoberta de conhecimento em bases de dados*. In: Anais do V Simpósio de Excelência em Gestão e Tecnologia-SEGeT. p. 4-5.
- Buntine, W. L. (1994). Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, p. 159–225.
- Caldeira, C. P. (2004). *Introdução ao modelo de dados relacional*. Departamento de Informática, Universidade de Évora.
- Camilo, C. O. (2009). *Mineração de Dados Conceitos, Tarefas, Métodos e Ferramentas*. Universidade Federal de Goiás, p. 1-29.
- Carvalho, L. A. V. (2005). *Datamining: A Mineração de Dados, no Marketing, Medicina, Economia, Engenharia e Administração*. Rio de Janeiro: Editora Ciência Moderna Ltda.
- Cavalcante, R. S. (2014). *Descoberta de conhecimento na plataforma lattes: um estudo de caso no instituto federal de Goiás*. Engenharia de Produção e Sistemas, Pontifícia Universidade Católica de Goiás, Goiânia. Tese de Mestrado.
- Chapman, P., Clinton, J., & et al. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. DaimlerChrysler.
- Côrtes, S. C., Porcaro, R. M., & Lifschitz, S. (2002). *Mineração de Dados-Funcionalidades, Técnicas e Abordagens*. Pontifícia Universidade Católica.

- Cortez, P. A. R. (2002). *Modelos inspirados na natureza para a previsão de séries temporais*. Universidade de Minho, Portugal.
- Dantas, E. R., Júnior, J. C., Lima, D. S., & Azevedo, R. R. (2008). *O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões*. V Simpósio de Excelência em Gestão e Tecnologia.
- Domingues, M. A. (2004). *Generalização de regras de associação*. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos. Dissertação de Mestrado.
- European Pet Food Industry Federation, FEDIAF. (2021). *Estatísticas europeias*. Acesso em: 23 de março de 2023, Disponível em: <https://europeanpetfood.org/about/statistics/>.
- Faraco, C. B. & Soares, G. M. (2013). *Fundamentos do comportamento animal canino e felino*. Editora Medvet, São Paulo. 242p. 1ª ed.
- Fayyad, U., Piatetsky-Shapiro, G., & Smith, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17. [www.ffly.com/](http://www.ffly.com/)
- Goldberger, J., Roweis, S., Hinton, G., & Salakhutdinov, R. (2004). *Neighbourhood Components Analysis*. Department of Computer Science, University of Toronto.
- Gomes, B. M. V. (2011). *Previsão de Churn em Companhias de Seguros*. Curso de Escola de Engenharia, Departamento de Informática, Universidade do Minho, Braga. Dissertação de Mestrado.
- Haykin, S. (1998). *Neurais networks: a comprehensive foundation*. Prentice Hall PTR.
- Heckerman, D., Geiger, D., & Chickering, D. M. (1995). *Learning Bayesian Networks: The Combination of Knowledge and Statistical*. *Machine Learning*, vol. 20, p. 197-243.
- Machado, D. D. S., Maciel, T. T., Machado, J. C., & Prezoto, H. H. (2017). *Interação entre gatos domésticos (Felis silvestris catus Linnaeus, 1758) cativos e seres humanos*. *Revista Brasileira de Zoociências*, 67–72.
- Nogueira, D. (2014). *Agile Data Mining: Uma metodologia ágil para o desenvolvimento de projetos de data mining*. Curso de Mestrado Integrado em Engenharia Informática e Computação, Faculdade de Engenharia da Universidade do Porto, Portugal. Dissertação de Mestrado.

- Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.
- Podberscek, A. L. (1997). *Illuminating issues of companion animal welfare through research into human–animal interactions*. *Animal Welfare*, v.6, p.365-372.
- Quinlan, J. R. (1986). *Induction of Decision Trees*. In *Machine Learning*, vol. 1, no. 1, pp. 81-106.
- Quinlan, J. R. (1998). *Programs for Machine Learning*, Morgan Kaufmann Publishers. Inc. 1998.
- Quintela, H., Santos, M. F., & Cruz, P. (2005). *Sistemas de Conhecimentos Baseados em Data Mining - Aplicação à Análise da Estabilidade de Estruturas Metálicas*. Universidade do Minho - Escola de Engenharia. Dissertação de Mestrado
- Rochlitz, I. (2000). *Feline welfare issues*. In: Turner, D.C.; Bateson, P. *The domestic cat: The Biology of its Behaviour*. Cambridge: Cambridge University Press, p. 207-226.
- Serpell, J. (1996). *Evidence for an association between pet behavior and owner attachment levels*. School of Veterinary Medicine, University of Pennsylvania.
- Serpell, J.A. (2000). *Domestication and history of the cat*. In: Turner, D.C.; Bateson, P. *The domestic cat: The Biology of its Behaviour*. Cambridge: Cambridge University Press, p. 180-191.
- Sistema de Informação de Animais de Companhia – SIAC (2021). *O registo de animais de companhia promove o bem-estar animal*. Acesso em: 28 de março de 2023, Disponível em: <https://www.siac.vet/>.
- Souza, M. S. R. (1998). *Mineração de Dados: Uma implementação fortemente acoplada a um sistema gerenciador de banco de dados paralelo*. Universidade Federal do Rio de Janeiro – UFRJ, Dissertação de Mestrado.
- Sparkes, A., & Ellis, S. (2016). Guia ISFM para estresse e saúde felina. *Gerenciando emoções negativas para melhorar a saúde e o bem-estar felino*. Cuidados Internacionais com Gatos.
- Stammbach, K. B., & Turner, D.C (1999). *Understanding the Human–Cat Relationship: Human Social Support or Attachment*. *Anthrozöos*, v.12, p.162-168.

- Takahashi, T. (2000). *Sociedade da informação no Brasil: livro verde*. Ministério da Ciência e Tecnologia.
- Thomé, A. C. G. (2008). *Redes neurais: uma ferramenta para KDD e data mining*.
- Toloo, M., Sohrabi, B., & Nalchigar, S. (2009). *A new method for ranking discovered rules from data mining by DEA*. *Expert Systems with Applications*, 36(4), 8503–8508. <https://doi.org/10.1016/j.eswa.2008.10.038>
- Witten, I. H., Frank, E., Hall, M. A. *Data Mining - Practical Machine Learning Tools and Techniques*, Elsevier, Burlington, MA 01803, USA, 2011.
- Zhang, H. (2004). *The Optimality of Naive Bayes*. In 2004 FLAIRS Conference – AAAI. [www.aaai.org](http://www.aaai.org)