

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS COMPUTACIONAIS
PARA O AGRONEGÓCIO

MILIANO DE BASTIANI

**APLICAÇÃO DE CONTROLE ESTATÍSTICO DE PROCESSOS E
ALGORITMOS DE MINERAÇÃO DE DADOS NA GESTÃO DA
PRODUÇÃO DE FRANGOS DE CORTE**

PROJETO DE DISSERTAÇÃO

MEDIANEIRA-PR

2018

MILIANO DE BASTIANI

**APLICAÇÃO DE CONTROLE ESTATÍSTICO DE PROCESSOS E
ALGORITMOS DE MINERAÇÃO DE DADOS NA GESTÃO DA
PRODUÇÃO DE FRANGOS DE CORTE**

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio – PPGTCA – da Universidade Tecnológica Federal do Paraná – UTFPR – Campus Medianeira, como requisito parcial para obtenção do título de Mestre em Tecnologias Computacionais para o Agronegócio / Área de Concentração: Tecnologias Computacionais Aplicadas à Produção Agrícola e Agroindústria.

Orientador(a): Prof. Dr. José Airton Azevedo dos Santos

Coorientador(a): Prof(a). Dr(a). Carla Adriana Pizarro Schmidt

Coorientador(a): Prof(a). Dr(a). Gloria Patricia López Sepúlveda

MEDIANEIRA-PR

2018

Dados Internacionais de Catalogação na Publicação

B326a Bastiani, Miliano de

Aplicação de controle estatístico de processos e algoritmos de mineração de dados na gestão da produção de frangos de corte / Miliano de Bastiani – 2018.
76 f. : il. ; 30 cm.

Orientador: José Airton Azevedo dos Santos.

Coorientadora: Carla Adriana Pizarro Schmidt.

Coorientadora: Gloria Patricia Lopez Sepulveda

Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio. Medianeira, 2018.

Inclui bibliografias.

1. Mineração de dados (Computação). 2. Agroindústria. 3. Controle de processo - Métodos estatísticos. 4. Tecnologias Computacionais - Dissertações. I. Santos, José Airton Azevedo dos. orient. II. Schmidt, Carla Adriana Pizarro, coorient. III. Sepulveda, Gloria Patricia Lopez, coorient. IV. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio..V. Título.

CDD: 004

Biblioteca Câmpus Medianeira
Marci Lucia Nicodem Fischborn 9/1219



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Medianeira
Diretoria de Pesquisa e Pós-Graduação
Mestrado em Tecnologias Computacionais para o Agronegócio



TERMO DE APROVAÇÃO

APLICAÇÃO DE CONTROLE ESTATÍSTICO DE PROCESSOS E ALGORITMOS DE MINERAÇÃO DE DADOS NA GESTÃO DA PRODUÇÃO DE FRANGOS DE CORTE

Por

MILIANO DE BASTIANI

Essa dissertação foi apresentada às quatorze horas, do dia dezoito de junho de dois mil e dezoito, como requisito parcial para a obtenção do título de Mestre em Tecnologias Computacionais para o Agronegócio, Linha de Pesquisa Tecnologias Computacionais Aplicadas à Agroindústria, no Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio - PPGTCA, da Universidade Tecnológica Federal do Paraná. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

Prof. Dr. José Airton Azevedo dos Santos (Orientador – PPGTCA)

Prof. Dr. Carlos Aparecido Fernandes (Membro Externo – UTFPR)

Prof. Dr. Jair Antonio Cruz Siqueira (Membro Externo – UNIOESTE, com participação à distância,

sendo a assinatura na presente ata dispensada conforme Regulamento da

Pós-Graduação *Stricto Sensu* da UTFPR)

A via original com as assinaturas encontra-se na secretaria do programa.

A Deus, a minha mãe, noiva e aos meus amigos *companheiros* de todas as horas e
que me incentivaram sempre.

AGRADECIMENTOS

Ao Prof. Dr. Orientador, braço amigo de todas as etapas deste trabalho, paciente, solícito e sempre disponível para ajudar no que for preciso.

A minha família, pela confiança, motivação e o apoio para conseguir superar todos os momentos difíceis e fazer dessa oportunidade a possibilidade de se tornar mestre.

Aos amigos e colegas, pela força, pela vibração em relação a esta jornada e compreensão pelos momentos de ausência dedicados para realizar esse trabalho.

Aos professores e colegas de Curso, pois juntos trilhamos uma etapa importante de nossas vidas.

Aos profissionais entrevistados, pela concessão de informações valiosas para a realização deste estudo.

A todos que, com boa intenção, colaboraram para a realização e finalização deste trabalho.

"Se existe uma forma
de fazer melhor, descubra-a."

Thomas Edison

RESUMO

BASTIANI, Miliano de. Aplicação de controle estatístico de processos e algoritmos de mineração de dados na gestão da produção de frangos de corte. 2018. 76f. Dissertação (Mestrado em tecnologias computacionais para o agronegócio), Universidade Tecnológica Federal do Paraná, Campus Medianeira, 2018.

Este trabalho tem por objetivo aplicar técnicas de controle estatístico de processo e algoritmos de mineração de dados na análise, previsão e classificação de variáveis produtivas que compõem as características das aves produzidas por uma cooperativa localizada na região oeste paranaense. A base de dados disponibilizada pela empresa apresenta um histórico de movimentação de 2 anos, contendo as principais variáveis de produção (Linhagem, Peso Médio, Índice de Eficiência Produtiva, Conversão Alimentar, Viabilidade, Idade do Lote e Ganho de Peso Diário), desde o alojamento até a fase de abate. Para analisar se os processos de produção de frangos de corte, das linhagens Coob e Coob Slow, atendiam as expectativas da cooperativa utilizou-se o peso médio de 186 lotes de aves. Verificou-se que os índices de capacidade estão abaixo de um, o que classifica os processos de produção dos aviários como incapazes. Três algoritmos de análise de dados (Multilayer Perceptron, Linear Regression e Naive Bayes), do software WEKA, foram utilizados na implementação dos modelos de previsão e classificação. Para esta implementação foram utilizados 6000 lotes de aves das linhagens Coob e Coob Slow, obtidos no período compreendido entre 01/07/2014 e 31/07/2016. Os resultados obtidos mostram que os modelos de previsão fornecem estimativas confiáveis para as variáveis de resposta: Peso Médio e Índice de Eficiência Produtiva e o modelo de classificação apresenta um bom desempenho na classificação das linhagens das aves.

Palavras-chave: mineração de dados; controle estatístico de processo; agroindústria.

ABSTRACT

BASTIANI, Miliano de. Application of statistical process control and data mining algorithms in the management of the production of broiler chickens. 2018. 76f. Dissertation (Master in Computational Technologies for Agribusiness), Federal Technological University of Paraná, Campus Medianeira, 2018.

This work aims to apply techniques of statistical process control and data mining algorithms in the analysis, prediction and classification of productive variables that make up the characteristics of the birds produced by a cooperative located in the western region of Paraná. The database provided by the company has a 2-year history of movement, containing the main production variables (Lineage, Average Weight, Productive Efficiency Index, Food Conversion, Feasibility, Lot Age and Daily Weight Gain), from the until the slaughter stage. In order to analyze whether the production processes of broilers, Coob and Coob Slow, met the cooperative's expectations, the average weight of 186 poultry lots was used. It has been found that the capacity indices are below one, which classifies the production processes of the aviaries as incapable. Three algorithms for data analysis (Multilayer Perceptron, Linear Regression and Naive Bayes) of WEKA software were used in the implementation of prediction and classification models. For this implementation, 6000 batches of birds of the Coob and Coob Slow strains were used, obtained between 01/07/2014 and 07/31/2016. The results show that the prediction models provide reliable estimates for the response variables: Average Weight and Productive Efficiency Index and the classification model shows a good performance in the classification of the bird strains.

Key-words: data mining; statistical process control; agroindustry.

LISTA DE FIGURAS

Figura 1 - Percentual da Produção Aves por Estado	22
Figura 2 - Estrutura da cadeia produtiva de frango de corte	25
Figura 3 - Etapas do processo de descoberta do conhecimento.....	27
Figura 4 - Neurônio Biológico	31
Figura 5 - Estrutura de um Neurônio.	32
Figura 6 - Arquitetura de uma rede MLP	33
Figura 7 - Carta de controle.....	37
Figura 8 - Zonas A, B e C do gráfico de controle	38
Figura 9 - Cartas de Controle X-AM.	40
Figura 10 - Região Oeste Paranaense.....	43
Figura 11 - Arquivo de entrada de dados.	47
Figura 12 - Seleção de atributos – CfsSubsetEval.	51
Figura 13 - Fluxograma de análise do índice de capacidade do processo de produção de frangos de corte.	52
Figura 14 - Outros padrões de causas especiais - Coob Slow.....	54
Figura 15 - Outros padrões de causas especiais - Coob	55
Figura 16 - Gráfico $\bar{X}_i$ - MR – Coob Slow – Peso Médio.....	55
Figura 17 - Gráfico $\bar{X}_i$ - MR – Coob – Peso Médio	56
Figura 18 - Gráfico de Probabilidade Normal – Coob Slow	57
Figura 19 - Gráfico de Probabilidade Normal – Coob.....	57
Figura 20 - Histograma e Índice de Capacidade C_{pk} – Coob Slow.....	58
Figura 21 - Histograma e Índice de Capacidade C_{pk} – Coob.....	58
Figura 22 - Redes MLP 5-3-1 e Rede MLP 3-2-1	61
Figura 23 - Dados preditos em função de dados observados	63
Figura 24 - Dados preditos em função de dados observados – 400 lotes.....	64

Figura 25 - Dados preditos em função de dados observados – MLP e Cross-validation65

LISTA DE TABELAS

Tabela 1 - Classificação do processo.....	41
Tabela 2 - Níveis de Concordância – Índice Kappa	50
Tabela 3 - Classificação dos atributos.....	59
Tabela 4 - Variáveis de entrada e saída para os algoritmos	60
Tabela 5 – Resultados da aplicação dos algoritmos MLP e LR	62
Tabela 6 - Resultados da aplicação do algoritmo MLP – Cross-validation e Holdhout.	65
Tabela 7 - Resultados da aplicação do algoritmo Naive Bayes.....	67

LISTA DE SIGLAS

ABPA	Associação Brasileira de Proteína Animal
AM	Amplitude Móvel
ARFF	Attribute Relation File Format
ART	Teoria da Ressonância Adaptativa
BAM	Memória Bidirecional Associativa
BSB	Brain State in a Box
CA	Conversão Alimentar
CB	Coob
CEP	Controle Estatístico de Processo
CF	Coob Fast
C_{PK}	Índice de Capacidade de Processo
CS	Coob Slow
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
GPD	Ganho de Peso Diário
IAC	Ativação de Interação e Competição
IEP	Índice de Eficiência Produtiva
LC	Linha Central
LIC	Linha Inferior de Controle
LM	Linha Média
LR	Linear Regression
LSC	Linha Superior de Controle
LTS	Limite de Tolerância Superior
MR	Amplitude Móvel
MLP	Perceptron Multi-Camadas

PM	Peso Médio
PNN	Rede Neural Probabilística
RBF	Radial Bases Function
RNA	Rede Neural Artificial
RO	ROSS
SVM	Suport Vector Machines
UBA	União Brasileira de Avicultura
VIAB	Viabilidade
X_i	Valores Individuais

SUMÁRIO

1 INTRODUÇÃO	17
2 OBJETIVOS.....	19
2.1 OBJETIVO GERAL	19
2.2 OBJETIVOS ESPECÍFICOS	19
3 FUNDAMENTAÇÃO TEÓRICA.....	20
3.1 A ORIGEM DA AVICULTURA NO BRASIL.....	20
3.2 A ORIGEM DA AVICULTURA NO ESTADO DO PARANÁ.....	22
3.3 LINHAGEM DE AVES	23
3.4 CADEIA PRODUTIVA DE FRANGO.....	25
3.5 DESCOBERTA DE CONHECIMENTO	27
3.5.1 Tarefas da mineração de dados.....	29
3.5.2 Métodos de mineração de dados	30
3.6 UTILIZAÇÃO DE MINERAÇÃO DE DADOS NA AVICULTURA	36
3.7 CAPACIDADE DO PROCESSO DE PRODUÇÃO.....	37
4 PROCEDIMENTOS METODOLÓGICOS.....	43
4.1 A EMPRESA	43
4.2 TIPO DE PESQUISA.....	44
4.3 MINERAÇÃO DE DADOS	45
4.4 VARIÁVEIS PRODUTIVAS	45
4.5 O SOFTWARE WEKA.....	47
4.6 ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS.....	48
4.7 ÍNDICE DE CAPACIDADE	52
4.8 PESOS DAS AVES	54
5 ANÁLISE DE RESULTADOS	55

5.1 CAPACIDADE DO PROCESSO DE PRODUÇÃO	55
5.2 MODELOS DE PREVISÃO	60
5.3 ALGORITMOS	61
5.4 MODELO DE CLASSIFICAÇÃO	67
6 CONCLUSÃO	69
REFERÊNCIAS.....	71

1 INTRODUÇÃO

A avicultura brasileira é considerada como uma das mais estruturadas do mundo, com números de produtividade significativos, graças a programas de qualidade como o de genética, nutrição, manejo, biossegurança, boas práticas de produção, rastreabilidade, bem-estar animal e de preservação do meio ambiente. A avicultura está presente em todas as regiões do país, gerando um efeito positivo na economia da maioria dos estados brasileiros (União Brasileira de Avicultura - UBA, 2008).

O mercado mundial de carne de frango tem se caracterizado pelo aumento do consumo de frango, em função do crescimento da demanda da população, gerando assim o aumento da produtividade e a redução dos preços (Associação Brasileira de Proteína Animal - ABPA, 2015).

Segundo Oliveira e Nääs (2012), o aumento da produtividade está relacionado com vários fatores, como: melhoramento de linhagens e insumos, investimentos na automatização do processo de produção, aperfeiçoamento dos funcionários em relação ao manejo, controle de condições sanitárias, etc.

Os consumidores, devido ao aumento do consumo de carne de frango, tornaram-se bem mais seletivo e exigentes, gerando assim uma maior competitividade entre as empresas alimentícias, levando-as a procurar vantagens competitivas. Algo fundamental para obtenção dessas vantagens competitivas é a melhoria dos processos produtivos. Esse tipo de melhoria pode ser alcançado de forma efetiva mediante a utilização de métodos estatísticos e ferramentas da qualidade. Tais métodos e ferramentas permitem que os dados gerados pelo processo possam ser interpretados e analisados (RAMOS, 2013).

Segundo Montgomery (2016) o controle estatístico de processo (CEP) é uma ferramenta que traz junto consigo uma série de outras ferramentas da qualidade como: o fluxograma para o mapeamento e entendimento do processo, histograma, diagrama de causa e efeito, gráfico de Pareto que auxilia na identificação de fontes de causas

especiais e as cartas de controle que permitem a visibilidade de pontos fora do padrão desejado.

As cartas de controle são uma das principais ferramentas do controle estatístico de processo. Estas cartas, apesar de simples, são ferramentas fundamentais para auxiliar no monitoramento do processo produtivo, pois apontam a ocorrência ou tendência à falta de controle. Porém, mesmo um processo sob controle pode produzir itens não conformes. Portanto, não é somente suficiente ter um processo sob controle. É importante verificar se o processo é capaz de atender as especificações do setor de engenharia da indústria. Esta verificação pode ser realizada por meio da análise de índices de capacidade de processo (RIBEIRO; CATEN, 2012).

A tomada de decisão, no meio industrial, é outro importante fator para obtenção de vantagens competitivas. Segundo Azarias, Matos e Scandelari (2009), o processamento e a interpretação de informações, geradas por banco de dados, está entre os requisitos essenciais para uma boa tomada de decisão.

A mineração de dados é um processo analítico utilizado na exploração de grandes quantidades de dados. Buscando padrões consistentes e/ou relações sistemáticas entre as variáveis e, assim, validá-los aplicando os padrões detectados a novos subconjunto de dados. Basicamente as etapas de um processo de mineração de dados consiste na: exploração, construção do modelo e validação/verificação (GOLDSCHMIDT, PASSOS e BEZERRA, 2015).

Neste contexto, este trabalho tem por objetivo aplicar técnicas de controle estatístico de processo e algoritmos de mineração de dados na análise, previsão e classificação de variáveis produtivas que compõem as características das aves produzidas por uma cooperativa localizada na região oeste paranaense.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Este trabalho tem por objetivo aplicar técnicas de controle estatístico de processo e algoritmos de mineração de dados na análise, previsão e classificação de variáveis produtivas que compõem as características das aves produzidas por uma cooperativa localizada na região oeste paranaense.

2.2 OBJETIVOS ESPECÍFICOS

- a) Realizar a limpeza e o tratamento dos dados.
- b) Obter informações sobre a capacidade do processo de produção de frangos de corte por meio do índice de capacidade C_{pk} .
- c) Implementar modelos de regressão e classificação, no software WEKA, para previsão e classificação de índices zootécnicos obtidos do banco de dados fornecido pela empresa.

3 FUNDAMENTAÇÃO TEÓRICA

A presente seção inicia com um breve embasamento inicial, apresentando conceitos sobre os métodos de produção de frangos de corte, tópicos sobre mineração de dados e controle estatístico de processos. Destaca-se também, nesta seção, um tópico sobre técnicas de mineração de dados utilizadas neste trabalho.

3.1 A ORIGEM DA AVICULTURA NO BRASIL

De acordo com o relatório da ABPA (2015), a chegada da avicultura no Brasil, está relacionada ao descobrimento do nosso país pelos portugueses. Por volta de 1502, a comitiva de Pero Vaz de Caminha, escrivão de Pedro Álvares Cabral, atracou no litoral sul do atual estado da Bahia trazendo, a bordo das caravelas, as primeiras matrizes. Essas matrizes, por sua vez tinham uma origem peculiar, eram originadas do cruzamento realizado entre diversas raças, sendo criadas nos quintais das casas.

No início, a ausência de dificuldades na criação de aves fez com que a avicultura se desenvolvesse de forma artesanal nas regiões litorâneas, onde as raças mestiças eram comercializadas vivas em feiras e comércios locais. A partir de 1895, foram realizados estudos pela empresa Leste Basse-Cour, no estado do Rio de Janeiro, na tentativa de fundar os primeiros aviários utilizando raças puras, chamados de basse-cour, servindo de grande impulso para alavancar a avicultura nacional (ESPÍNDOLA, 2012).

Para Albino e Tavernari (2008), a ascensão definitiva para a criação de frangos de corte, aconteceu a partir de década de 1950 com a expansão de grandes pólos agropecuários em nosso país, presentes nos estados de São Paulo e Rio de Janeiro. Com o intuito de fortalecer a avicultura em nosso país, foi realizada a introdução de novas linhagens, conhecidas como *Leghorn* e *New Hampshire*, visando

fazer a substituição das raças rústicas que eram oferecidas em feiras locais da época, por espécies com melhores resultados no processo de produção avícola.

Além da utilização de novas raças, era necessário ampliar as instalações para novas plantas produtivas. Isso aconteceu entre as décadas de 1970 à 1980, em que novas empresas avícolas foram abertas com o objetivo de centralizar o capital no setor, através de investimentos em inovações tecnológicas, utilização de novas linhagens de matrizes e modernos equipamentos em todas as etapas: de criação, abate e processamento de frangos de corte. Por isso, nesse mesmo período novos frigoríficos foram instalados nos estados de São Paulo, Minas Gerais, Paraná e Rio Grande do Sul. O Governo Federal, também teve uma contribuição muito importante com a criação da Empresa Brasileira de Pesquisa Agropecuária/aves reforçando o empenho nacional na criação de um sistema de inovação independente na área de produção genética avícola (ABPA, 2015).

Para Espíndola (2012), nesse processo de expansão do setor da avicultura no Brasil, algumas empresas foram de suma importância e tiveram destaques fundamentais. A partir de 1970, a Sadia, uma das empresas pioneiras no setor, foi responsável por estabelecer na região oeste do estado de Santa Catarina, o modelo de produção de aves integrada, oriundo de pesquisas realizadas nos Estados Unidos e trazidas para o Brasil. Este modelo foi adotado posteriormente por várias empresas (Perdigão, Seara, entre outras).

Com a produção avançando em grande escala, após a década de 1990, aconteceu a abertura da economia e comercialização de produtos agroindustriais para o mercado latino-americano. Para Caldarelli e Camara (2013), a expansão na economia dos setores agroindustriais expôs a produção brasileira a nível de concorrência mundial, obrigando as indústrias realizarem a reestruturação das suas estratégias competitivas de mercado, bem como reorganizar os processos tecnológicos envolvidos na produção da cadeia produtiva de frangos de corte. O processo de reorganização da cadeia produtiva, beneficiou principalmente a região Centro-Oeste através da implantação de novos projetos, de plantas industriais, localizados nos estados de Mato Grosso, Rondônia, Acre, Tocantins (BARCZCZ e LIMA FILHO 2009).

De acordo com Albino e Tavernari (2008), a expansão produtiva transformou a região Centro-Oeste em um pólo de exploração para as indústrias processadoras da carne de frango. Isso se deve aos seguintes fatos: os produtores desta região trabalhavam com um número reduzido de granjas, proporcionando maior capacidade de produção, utilizavam raças puras de alto desempenho produtivo e a região ser a maior produtora de milho e soja, produtos esses usados para a base de alimentação das aves.

3.2 A ORIGEM DA AVICULTURA NO ESTADO DO PARANÁ

Com o avanço do processo de transnacionalização do capital agrícola, a avicultura industrial brasileira foi disseminada para outras regiões do país, dando início as atividades na região sul. A evolução da avicultura, foi influenciada diretamente pelas demandas comerciais e produtivas originadas na região, devido a instalação de agroindústrias do setor. Foi a partir do final da década de 1970, que a avicultura teve um papel muito importante para o desenvolvimento do agronegócio no estado do Paraná (RIPPEL, 2005).

As cooperativas agrícolas, em busca de terras férteis e com condições favoráveis para o desenvolvimento do processo de produção de aves, ocuparam, a partir de 1980, a mesoregião oeste do estado do Paraná. Estas cooperativas passaram a atuar na cadeia produtiva de frangos de corte, tendo como principal objetivo implementar inovações comerciais e produtivas no setor (BELUSSO e HESPANHOL, 2010).

Para Caldarelli e Camara (2013), nas últimas décadas o setor da avicultura de corte do estado teve um aumento significativo na produção, elevando o estado do Paraná a um nível de grande produtor.

O estado do Paraná tem se destacado fortemente na produção avícola, sendo o principal produtor de frangos de corte do Brasil, apresentando números expressivos,

como pode ser visto na Figura 1. Esses resultados, são possíveis devido a constante evolução no processo de produção, o aperfeiçoamento de novos métodos tecnológicos, durante as etapas da cadeia produtiva, e a constantes estudos relacionados ao melhoramento genético das linhas de aves (ABPA, 2015).

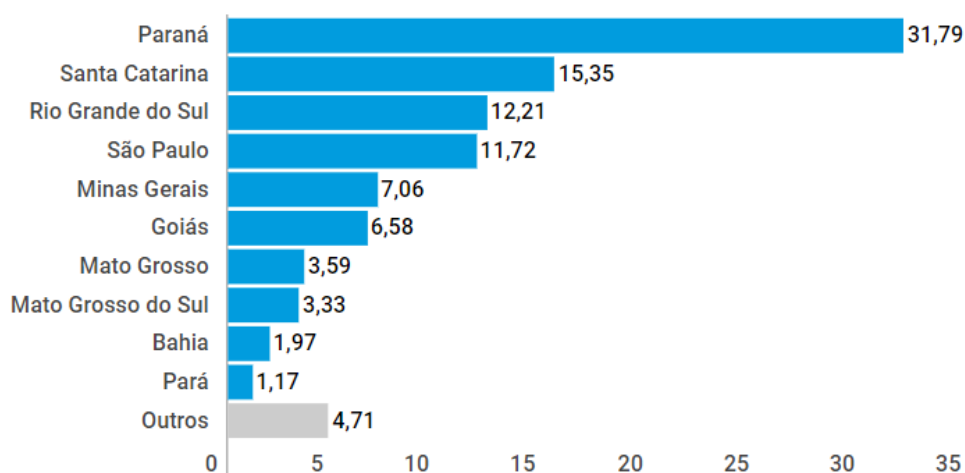


Figura 1 - Percentual da Produção Aves por Estado

Fonte: EMBRAPA (2017).

3.3 LINHAGEM DE AVES

Um dos fatores que contribuiu para a evolução da cadeia produtiva, está relacionado ao avanço genético das linhagens das aves no decorrer dos anos. No início, as aves eram mestiças, de raça comum e eram criadas de forma livre, soltas nos quintais das casas, alimentando-se apenas de resíduos alimentares. Foi a partir dos anos de 1920, com o início da implantação do processo de produção de plantéis avícolas, que o primeiro presidente da Sociedade Brasileira de Avicultura, Delgado de Carvalho, resolveu implantar no Brasil a raça *Orpington*, sendo importada da Inglaterra e avaliada como a melhor galinha do mundo, por possuir características consideradas positivas como: boa poedeira e excelente produtora de carne (ABPA, 2015).

Para Espíndola (2012), outras raças passaram a ser estudadas na empresa Leste Basse-Cour, criada no estado do Rio de Janeiro, com o objetivo de fazer o manejo de raças pertencente a outros países. Além da Inglaterra, foram trazidas dos Estado Unidos a raça *Plymouth Rock* após a visita realizada pelo Dr. Salvador de Mendonça, antigo ministro diplomático em Washington. Essa raça, constitui o tipo carijó, sendo a base de plantéis utilizados no meio rural, a qual os nossos antepassados tanto se orgulhavam.

Com o estudo realizado com novas espécies de aves, a produção de frango crescia rapidamente, graças aos pioneiros como Delgado de Carvalho e J. Wilson da Costa, que estudaram técnicas de manejo europeias e norte-americanas, para lançar dessa forma as bases de uma verdadeira avicultura industrial. Com o crescimento, a carne de frango já estava presente na dieta alimentar dos brasileiros, só que as aves ainda demoravam aproximadamente seis meses para atingir um peso médio de abate de 2,5 quilos (BARCZCZ e LIMA FILHO 2009).

Para Albino e Tavernari (2008), diversos estudos foram realizados na tentativa de trazer, para o nosso país, novos métodos e raças, criados em outros centros tecnológicos, com o objetivo de auxiliar na evolução da produção avícola brasileira. Somente a partir da década de 1930, no continente europeu e na América do Norte, foram realizadas as primeiras experiências positivas referentes a sexagem de pintainhos pela cor e a descoberta do valor nutritivo dos alimentos ingeridos pelas aves em comparação com o seu desempenho. Dava-se início, ao processo de estudo referente a evolução genética em busca de raças puras que pudessem proporcionar um aumento de rendimento em todas as etapas do processo produtivo das aves.

A EMBRAPA (2017) avalia que um número superior a 300 raças puras de aves e suas variedades já foram desenvolvidas até o momento, sendo que um valor bem expressivo possui representação comercial. Algumas das raças mais antigas ainda hoje são mantidas em estações relacionadas a experimentos genéticos. Por isso, deve-se ter ciência de que o banco genético das linhagens serve como combustível para os cruzamentos que dão origem a diversidade das raças híbridas modernas. Em frangos de corte as linhagens mais utilizadas são: *Coob* (CB), *Coob Fast* (CF), *Coob Slow* (CS) e *Roos* (RO).

O estudo referente as linhagens, foi de extrema importância para a modernização da cadeia produtiva de frangos de corte, pois fez com que a avicultura se torna-se um dos primeiros setores do agronegócio (ESPÍNDOLA 2012).

3.4 CADEIA PRODUTIVA DE FRANGO

Nas últimas décadas, a cadeia produtiva de frango de corte está em evidência, no agronegócio brasileiro, devido a uma diversidade de recursos tecnológicos utilizados e a eficácia no seu gerenciamento. O Brasil possui um complexo agroindustrial, com potencial competitivo, com grande atuação no mercado internacional (CALDARELLI e CAMARA 2013).

Dessa forma, para Lima et al. (2012) a avicultura incentiva o emprego no campo para milhares de produtores que provêm a matéria-prima para a agroindústria. Portanto, ela reflete uma atividade de grande vigor e extrema importância no setor econômico-social. O crescimento da produção e do consumo da carne de frango tem inspirado outros setores responsáveis pelo fornecimento de matérias primas, como o da indústria genética, alimentação animal, sanidade animal e máquinas e equipamentos.

A cadeia produtiva da carne de frango envolve os segmentos relacionados à seleção genética, incluindo a criação de matrizes avós, chamadas de avozeiros, seguido da criação de matrizes pais, tecnicamente referenciado como matrizeiros, do período de crescimento e engorda realizado no aviário e das etapas realizadas pela agroindústria, como: o abate, a distribuição das aves, o fornecimento de rações e medicamentos, além de todo suporte e acompanhamento (ESPÍNDOLA, 2012).

Voilà e Triches (2012) definem a cadeia produtiva conforme apresentado na Figura 3.2. De acordo com os autores, as principais ligações da cadeia estão relacionadas a alguns termos técnicos utilizados nas fases da produção de frangos de corte como o avozeiro, matrizeiro, incubatório, aviário, frigorífico, varejo e consumidor

final e aos elos auxiliares que tem como objetivo prover insumos e outras demandas à cadeia principal como: a pesquisa e desenvolvimento genético, medicamentos, milho, soja e outras matérias-primas.

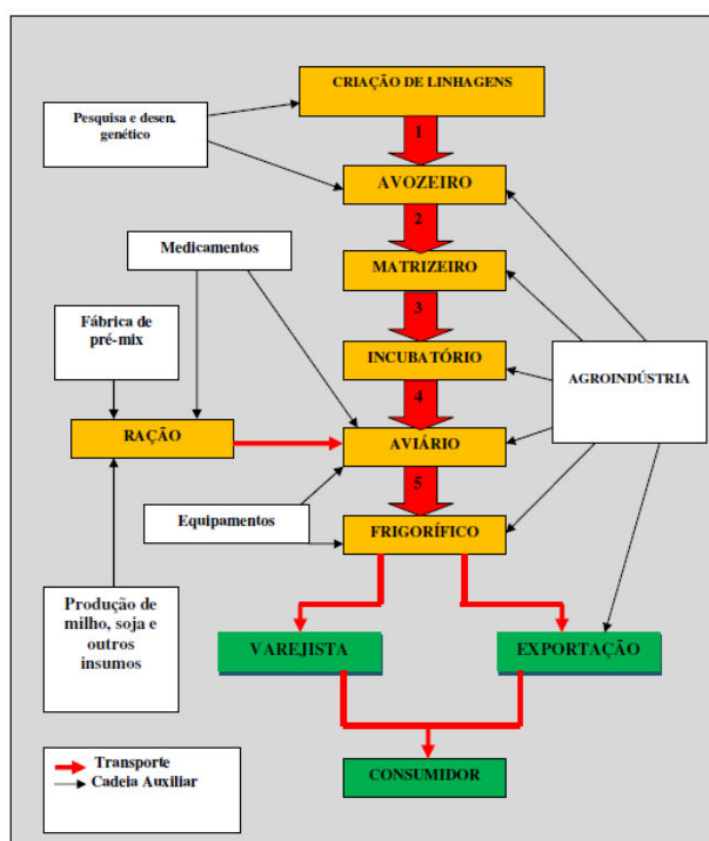


Figura 2 - Estrutura da cadeia produtiva de frango de corte

Fonte: Voilà; Triches (2012).

Uma seleção genética é necessária para aperfeiçoar a linhagem das aves que serão produzidas. Este processo é executado fora do Brasil. O avozeiro é onde ficam as poedeiras avós, que tem origem de ovos importados de linhagens específicas. O cruzamento entre avós dá origem às matrizes. Na grande maioria, os avozeiros são granjas de propriedades de empresas multinacionais (ALBINO e TAVERNARI, 2008).

Segundo Voilà e Triches (2012), a segunda ligação da cadeia é o matrizeiro, pois é nele onde ficam as matrizes que geram os ovos que darão origem aos frangos de corte. Este segmento é de responsabilidade da agroindústria. A próxima etapa, o

incubatório, é responsável por receber os ovos e chocá-los, sendo também de responsabilidade da agroindústria. Na sequência, os pintainhos são alojados nos aviários dos produtores, através de contratos realizados por meio da integração com os frigoríficos. Sendo essa etapa responsável pelo crescimento e engorda do frango, onde permanecem até completarem a idade ideal para serem abatidos. Após, as aves são destinadas ao frigorífico, sendo este responsável por executar o processo de abate e processamento da carne do frango. Na sequência, a carne é distribuída para os mercados varejistas, que tem a função de disponibilizar o produto para os consumidores.

3.5 DESCOBERTA DE CONHECIMENTO

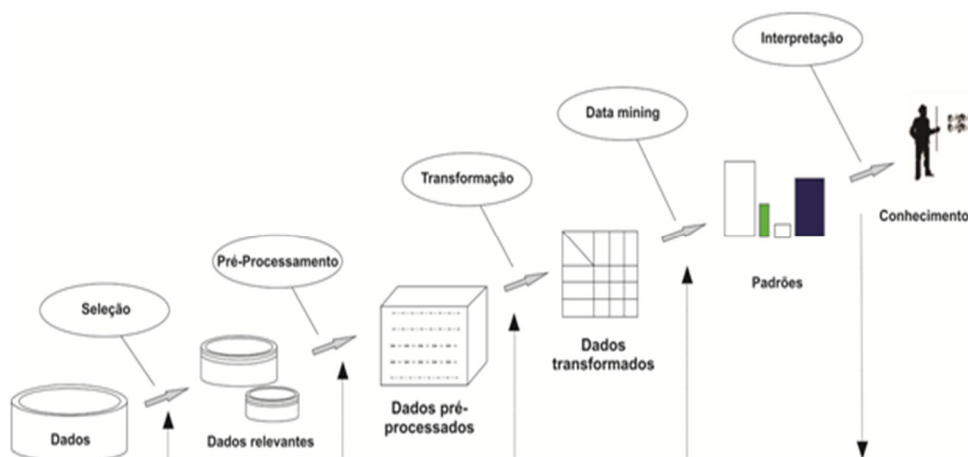
De acordo com Berthold (2010), a mineração de dados surgiu a partir da década 80 como um novo ramo da computação, devido à demanda das empresas em descobrir padrões para predizer e correlacionar dados. A mineração pode dar suporte a tomada de decisão, extraindo informação dos grandes volumes de dados que eram somente armazenados.

Segundo Goldschmidt, Passos e Bezerra (2015), a descoberta de conhecimento em bases de dados, do inglês *knowledge-discovery in databases* (KDD), pode ser compreendida como a forma de realizar o processo de extração de conhecimento oculto a partir de dados armazenados em um banco de dados.

Inicialmente, considerada como um sinônimo de KDD, a expressão Mineração de Dados é apenas uma das etapas da descoberta de conhecimento através do processo global do KDD. O conhecimento adquirido por meio da mineração de dados tem sido muito útil nas mais variadas áreas, como finanças, marketing, comércio, telecomunicações, meteorologia, agropecuária, bioinformáticas, medicina, entre outras (GOLDSCHMIDT, PASSOS e BEZERRA, 2015).

As etapas do processo de descoberta do conhecimento são apresentadas na

Figura 3.

**Figura 3 - Etapas do processo de descoberta do conhecimento****Fonte: FAYYAD et al. (1996).**

Segundo Fayyad et al. (1996), a mineração de dados não é um processo trivial, consiste em identificar padrões válidos nos dados, isto envolve métodos estatísticos, algumas ferramentas de visualização e a utilização de técnicas de inteligência artificial.

Já Zhou (2003) faz uma análise comparativa sobre três áreas consideradas como de maior significado dentro Mineração de dados: A estatística, o aprendizado de máquina e os bancos de dados.

Hand, Mannila e Smyth (2001), através do foco da estatística, apresenta o conceito de mineração de dados como uma análise de grandes estruturas de dados, com o objetivo de extrair os dados de uma forma que eles possam ser utilizados bem como compreendidos pelo proprietário dos dados.

Já o conceito de mineração de dados, por uma visão de banco de dados, é apresentada por Goldschmidt, Passos e Bezerra (2015), que relata como sendo um campo interdisciplinar que apresenta várias técnicas de reconhecimento de padrões, cálculos matemáticos, banco de dados e visualização do conjunto de informações que podem ser extraídos das grandes bases de dados.

Para Fayyad et. al. (1996), por uma perspectiva do aprendizado de máquina, a mineração de dados é apenas uma etapa no processo de descoberta de

conhecimento, que consiste em analisar os dados pela aplicação de algoritmos de descoberta, originando um conjunto de padrões de dados.

3.5.1 Tarefas da mineração de dados

Conforme o objetivo do estudo são definidas as tarefas e os algoritmos de mineração de dados que são utilizados para obter uma solução do problema apresentado (Matos, Chalmeta e Coltell, 2006). As técnicas de mineração de dados podem ser aplicadas a tarefas como classificação, estimativa, associação, segmentação e sumarização, sendo descritas, a seguir, as tarefas aplicadas neste trabalho.

a) Classificação

A Classificação, uma das tarefas mais comuns em mineração, visa identificar a qual classe um registro pertence, por meio de uma busca supervisionada dos dados. A tarefa de classificação utiliza o aprendizado supervisionado. Isto é, o modelo analisa um conjunto de registros contendo a indicação de sua classe, com o intuito de "aprender" como classificar um novo registro (CAMILO; SILVA, 2009).

b) Regressão

Para Goldschmidt, Passos e Bezerra (2015), a regressão é muito similar à

classificação, porém ela lida com registros numéricos, enquanto que a classificação lida com registros categóricos. Desse modo, por regressão, pode-se estimar o valor de uma determinada variável analisando-se os valores das demais.

3.5.2 Métodos de mineração de dados

Na mineração de dados os métodos são divididos em aprendizado supervisionado (preditivo) e não-supervisionado (descritivo). A principal diferença entre os métodos está no fato de que os métodos não-supervisionados podem ser executados com ausência da necessidade de se utilizar um atributo alvo, usando alguma medida de similaridade entre os atributos. Tem-se como exemplo de tarefas de mineração de dados não-supervisionadas o agrupamento e a associação. Já no aprendizado supervisionado, os métodos são providos através de um conjunto de dados que possuem uma variável alvo previamente definida, sendo os registros categorizados em relação a ela. As tarefas mais comuns de aprendizado supervisionado são as de classificação (MCCUE, 2015).

Para Camilo e Silva (2009), os algoritmos mais utilizados na mineração de dados são: árvores de decisão, classificação bayesiana, classificação baseada em regras, redes neurais, SVM (*Support Vector Machines*), classificação por regras de associação, aprendizado tardio, algoritmo genético, conjuntos aproximados, conjuntos nebulosos, regressão linear, regressão não-linear, métodos de particionamento, métodos hierárquicos, métodos baseados em densidade, métodos baseado em grade e métodos baseado em modelos.

Na sequência, faz-se uma breve apresentação de três técnicas de mineração de dados utilizadas neste trabalho:

a) Redes Neurais

A inspiração para a criação de redes neurais artificiais surgiu da tentativa de compreender o cérebro humano. As redes neurais artificiais são muito parecidas com as redes biológicas, tendo estrutura, função, técnicas de processamento de dados e métodos de cálculo muito similares. As redes neurais artificiais podem apreender com os dados e resolver problemas complexos (CAMILO; SILVA, 2009).

A técnica de redes neurais é apropriada às seguintes tarefas: classificação, estimativa e segmentação.

Exemplos de redes neurais: Perceptron, Rede MLP, Redes de Kohonen, Rede Hopfield, Rede BAM, Redes ART, Rede IAC, Rede LVQ, Rede Counterpropagation, Rede RBF, Rede PNN, Rede Time Delay, Neocognitron, Rede BSB (HAYKIN, 2003)

Neurônio Biológico:

O cérebro humano é composto por um grande número neurônios, aproximadamente 10 bilhões, com milhares de conexões entre eles. Cada neurônio é uma célula especializada que pode criar, receber e propagar sinais eletroquímicos.

O neurônio biológico é composto por: corpo celular; dendritos; axônio e suas terminações (Figura 4). As informações de entrada de outros neurônios são recebidas pelos dendritos, que conduzem essas informações para o corpo celular. No corpo celular esta informação é processada e outro impulso flui para o axônio. Caso este esteja conectado com outro neurônio, a informação será repassada.

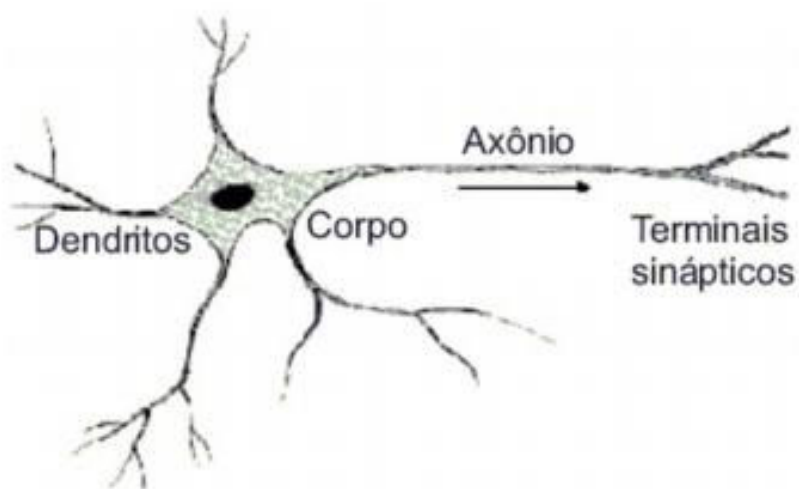


Figura 4 - Neurônio Biológico

Fonte: Goldschmidt, Passos e Bezerra (2015).

Neurônio Artificial:

O neurônio artificial também é composto por: dentritos (entradas), corpo celular (unidade de processamento) e terminais de axônio (saída).

Modelo Matemático Básico:

A estrutura de um neurônio artificial é apresentado na Figura 5.

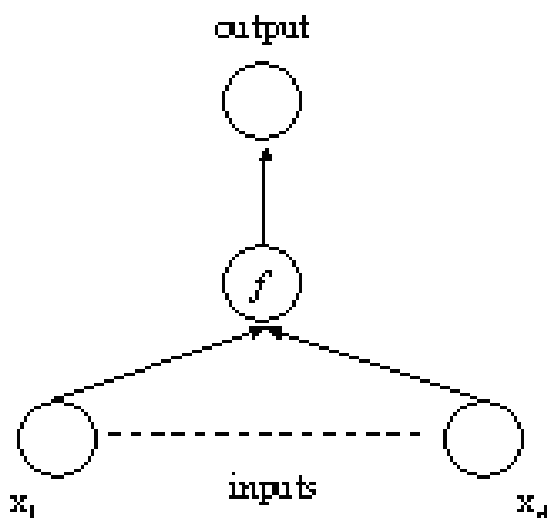


Figura 5 - Estrutura de um Neurônio.

Fonte: Han; Kamber (2011).

Os neurônios recebem sinais de muitas fontes. Estas fontes são referidas como variáveis de entrada (x_i), ou simplesmente entradas. As entradas são recebidas por conexões que possuem uma determinada “força”, chamada de peso (w_i). Ao receber os sinais uma soma ponderada (soma dos pesos) das entradas é formada, para compor a função de ativação f . A função de ativação do neurônio (Equação 1) é uma função matemática que converte a soma dos pesos dos sinais na saída do neurônio (SILVA, 2015).

$$f = (w_1x_1 + w_2x_2 \dots \dots + w_nx_n) \quad (1)$$

Pode-se usar como função de ativação (f) várias funções matemáticas. Dentre elas pode-se citar a: Identidade; Logística; Exponencial; Tangente hiperbólica; Sigmóide, etc.

Redes Multi-Layer Perceptron (MLP):

O treinamento de uma RNA é o processo de ajuste de seus pesos, através de um algoritmo de aprendizagem, que extrai características dos dados fornecidos e tem por objetivo gerar uma rede que desempenhe a tarefa de interesse (BINOTI, 2010). As redes treinadas, neste trabalho, são do tipo perceptron de múltiplas camadas (Figura 6), conhecidas como MLP (*Multilayer Perceptron*).

As redes MLP são utilizadas para resolver diversos tipos de problemas através de seu treinamento *backpropagation* (SILVA, 2015). São constituídas por uma camada entrada, uma ou mais camadas escondidas e uma camada de saída. São consideradas redes progressivas (*feedforward*). Nestas redes as saídas dos neurônios de uma determinada camada se conectam apenas às entradas dos neurônios da camada seguinte. As redes MLP apresentam uma função de ativação para cada neurônio na rede.

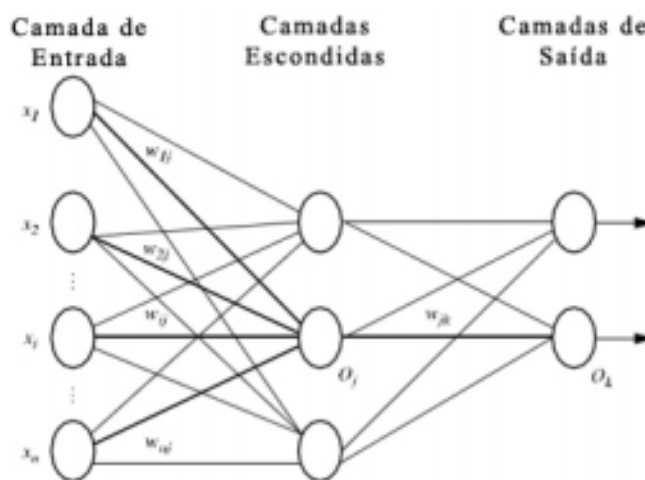


Figura 6 - Arquitetura de uma rede MLP

Fonte: Han; Kamber (2011).

Algoritmo Backpropagation:

O algoritmo *backpropagation* é o mais usado algoritmo de aprendizagem supervisionada, que busca minimizar o erro obtido e reajustar os pesos associados (SILVA, 2015). O objetivo é ajustar os pesos da rede, modificando-os até que a

aplicação de um conjunto de entradas produza a saída da rede que corresponda às saídas desejadas ou alvos.

b) Regressão Linear

Quando o resultado, ou classe, é numérica e todos os atributos são numéricos, a regressão linear é uma técnica a ser considerada. A idéia é expressar a classe como uma combinação linear de atributos, com peso pré-determinados:

$$y = w_0 + w_1a_1 + w_2a_2 + \dots + w_ka_k \quad (2)$$

Onde y é a classe, $a_1, a_2 + \dots + a_k$ são os valores dos atributos e $w_1, w_2 + \dots + w_k$ são os pesos. Os pesos são calculados através dos dados de treinamento. Em mineração de dados o algoritmo de regressão linear é muito utilizado para fazer previsão, e pode ser encontrado em vários softwares estatísticos.

c) Naive Bayes

O Classificador *Naive Bayes* é um método de aprendizagem probabilística baseado no Teorema de *Bayes*. Apesar de sua simplicidade, o classificador de *Naives Bayes* é muito utilizado, superando muitas vezes os métodos de classificação mais sofisticados. A classificação *baseyana* é um método de aprendizagem supervisionada, e também um método estatístico de classificação. O Classificador *Naive Bayes* aprende, a partir dos dados de treinamento, a probabilidade condicional de cada atributo dado o valor da classe.

A rapidez do algoritmo perfaz a sua maior vantagem e torna esta técnica de classificação mais rápida. A simplicidade do algoritmo torna-o adequado para conjuntos de dados com muitos atributos. A desvantagem desta técnica é que os atributos estão normalmente correlacionados e esse fato degrada a performance da

técnica a ponto de ser preterida por outras.

O algoritmo de *Naive Bayes* é baseado em probabilidade condicional. A probabilidade condicional é a probabilidade que um evento aconteça, devido a outro evento ter acontecido. A probabilidade que o Evento A ocorra, dado que o evento B aconteceu é representada por uma barra ($P(A|B)$). Ele usa a Regra de *Bayes*, uma fórmula que calcula a probabilidade por meio da contagem da frequência dos valores e de suas combinações nos dados históricos (Equação 3).

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (3)$$

onde A e B são eventos e $P(B) \neq 0$.

3.6 UTILIZAÇÃO DE MINERAÇÃO DE DADOS NA AVICULTURA

A aplicação de técnicas de mineração de dados, para obtenção de melhorias na avicultura, tornou-se freqüente nos últimos anos. Rodrigues, Vieira e Silva (2011), utilizaram técnicas de mineração de dados, em um abatedouro no estado de São Paulo, para avaliar as estimativas de mortalidade no pré-abate de frangos de corte. Já Lima e Rodrigues (2010) utilizaram mineração para identificar a importância do incubatório de ovos na avicultura de postura. A análise dos dados foi realizado com o Software SAS Enterprise Miner (SAS-EM) Versão 4.3, em uma base de dados de 10.300 registros. A identificação de fatores, por meio de mineração de dados, que influenciam os graus de lesões nas pernas das aves, durante o processo de produção, foi realizado por Cordeiro et al. (2012). Nesta pesquisa foi utilizado o software WEKA. Nascimento et al. (2011) utilizou mineração de dados na previsão do conforto térmico de frangos de corte. O experimento foi realizado em dois galpões avícolas, onde foram coletadas fotos termográficas, para montagem da base de dados.

3.7 CAPACIDADE DO PROCESSO DE PRODUÇÃO

A capacidade de processo, ou também expressa por capabilidade de processo, refere-se à como um processo estável, sem causas especiais, está agindo especificamente, isto é, se ele é capaz de produzir corretamente de acordo com que foi especificado pelo cliente/indústria (COSTA, EPPRECHT e CARPINETTI, 2010). Um processo instável, possui um comportamento imprevisível, portanto não tem sentido sua avaliação. Somente processos estáveis devem ter sua capacidade avaliada. As cartas de controle são ferramentas que informam sobre a estabilidade de um processo.

Causa especial é qualquer coisa que conduz uma observação além do limite de controle (MONTGOMERY, 2016). Também são aquelas cujas fontes de variações são relativamente grandes, bem maiores do que a variabilidade natural, sendo, porém, identificáveis e, ocorrem fora do sistema constante de variação. São evitáveis, e por gerarem defeitos, é necessário que sejam eliminadas. Um processo que esteja operando com a presença de causas especiais é dito fora de controle (RAMOS, 2013).

Cartas de Controle:

A carta de controle ou gráfico de controle tem como objetivo detectar desvios de parâmetros representativos do processo. Com sua utilização pode-se verificar a estabilidade de um processo (ALENCAR, LOPES, e SOUZA, 2007).

A carta de controle consiste na plotagem de amostras ordenadas no tempo, que são entendidas através de linhas horizontais, possuindo uma linha central (LC) ou também denominada linha média (LM), além de outras duas linhas chamadas de limite superior de controle (LSC) e limite inferior de controle (LIC), (Figura 7).

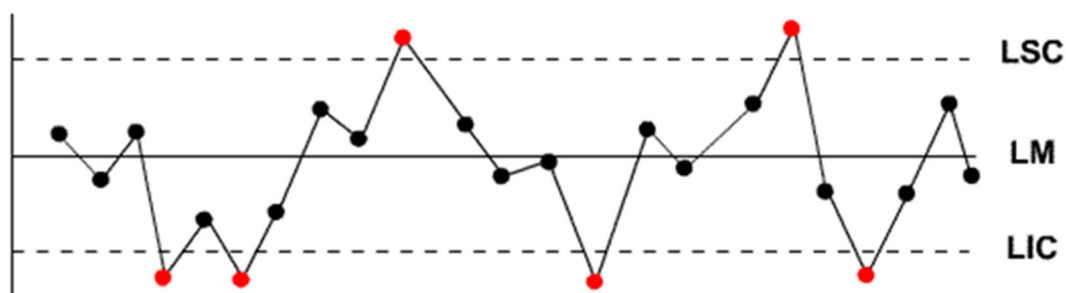


Figura 7 - Carta de controle

Fonte: Montgomery (2016).

As cartas de controle tem por objetivo (RAMOS, 2013):

- a) verificar se o processo em análise é estatisticamente estável, sem há presença de causas especiais;
- b) verificar se o processo continua estável, indicando quando se deve atuar sobre ele;
- c) permitir o aprimoramento constante do processo, por meio da redução de sua variabilidade.

Segundo Montgomery (2016), o processo está fora de controle se tiver:

- a) 1 ou mais pontos acima do LSC ou abaixo do LIC;
- b) 9 pontos consecutivos na zona C ou no mesmo lado do LC (Figura 8);
- c) 6 pontos consecutivos, todos aumentando ou todos diminuindo;
- d) 14 pontos consecutivos alternando para cima e para baixo;
- e) 2 de 3 pontos consecutivos na zona A ou além dela;
- f) 4 de 5 pontos consecutivos na zona B ou além dela;
- g) 15 pontos consecutivos na zona C (tanto acima quanto abaixo do LC);
- h) 8 pontos consecutivos na zona B.

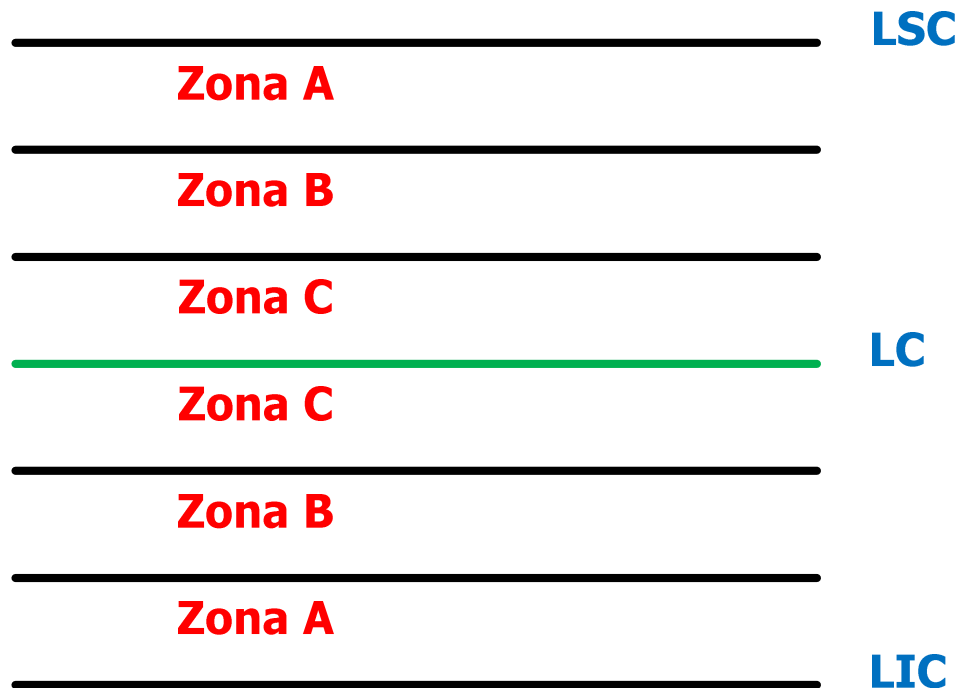


Figura 8 - Zonas A, B e C do gráfico de controle

Fonte: Montgomery (2016).

Cartas de Controle X-AM:

De acordo com Costa et al. (2010), as cartas X-AM (valores individuais e amplitude móvel – Figura 9), são utilizadas para monitorar o processo em casos em que as amostras são de tamanho $n = 1$. Utilizam-se estas cartas de controle para monitorar a estabilidade do processo ao longo do tempo para que seja possível identificar e corrigir as instabilidades de um processo. As equações para obter os parâmetros das cartas X-AM são apresentadas a seguir:

Gráfico X (Equações 4, 5 e 6):

$$LSC = \bar{x} + 3 \frac{\overline{AM}}{d_2} \quad (4)$$

$$LC = \bar{x} \quad (5)$$

$$LIC = \bar{x} - 3 \frac{\overline{AM}}{d_2} \quad (6)$$

Onde $\bar{x}$ é a média. A amplitude móvel ($\overline{AM}$) é dada por:

$$AM_i = |X_i - X_{i-1}| \quad i = 2, 3, \dots, m \quad (7)$$

$$\overline{AM} = \frac{\sum_{i=2}^m AM_i}{m-1} \quad (8)$$

Onde m é o número de elementos do conjunto de dados e X um elemento do conjunto.

Gráfico AM (Equações 9, 10 e 11):

$$LSC = D_4 \overline{AM} \quad (9)$$

$$LC = \overline{AM} \quad (10)$$

$$LIC = D_3 \overline{AM} \quad (11)$$

Onde d_2 , D_3 e D_4 são valores tabelados (MONTGOMERY, 2016).

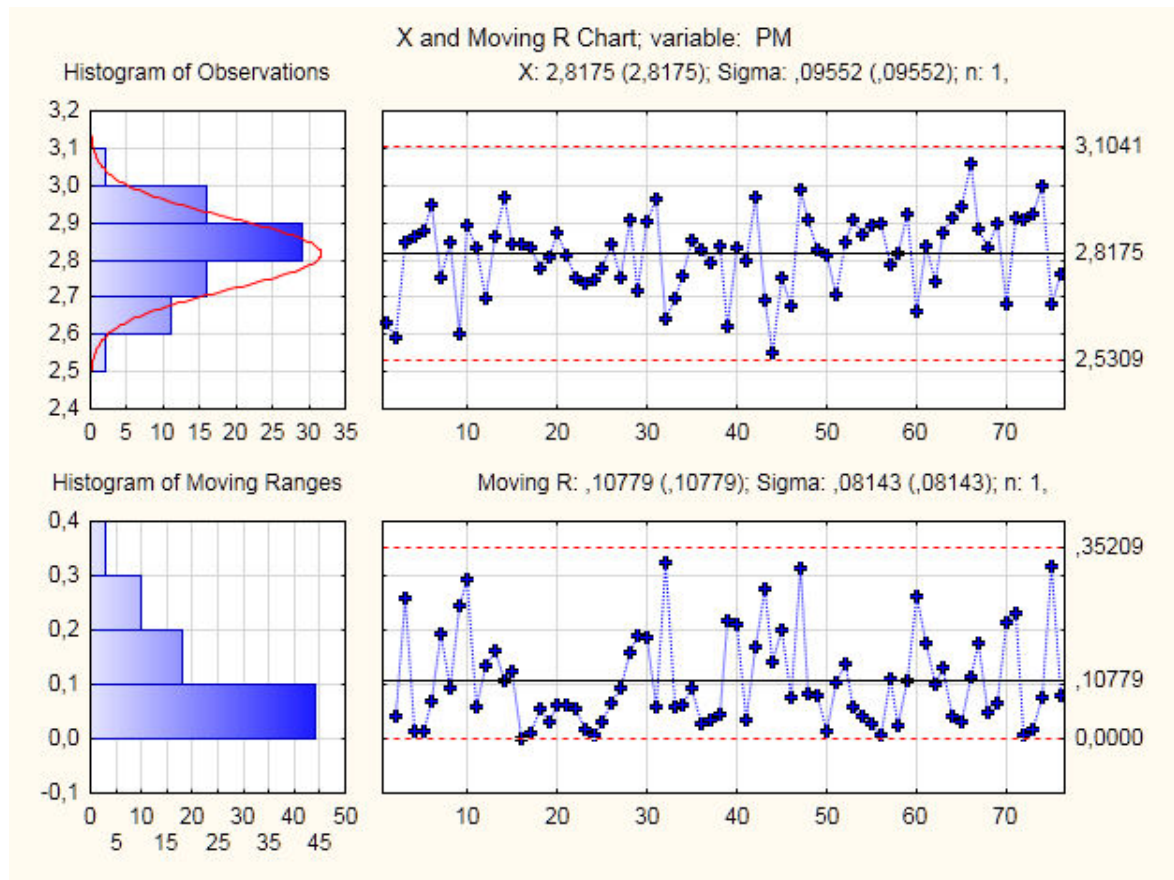


Figura 9 - Cartas de Controle X-AM

Fonte: Autoria Própria.

Índice de Capacidade C_{pk} :

O índice de capacidade do processo C_{pk} pode ser interpretado como uma medida da capacidade real do processo. Este índice é calculado por meio da Equação 12.

$$C_{PL} = \frac{(\mu - LTI)}{3\sigma}, \quad C_{PU} = \frac{(LTS - \mu)}{3\sigma}, \quad C_{pk} = \min(C_{PL}, C_{PU}) \quad (12)$$

Onde: LTS : Limite de tolerância superior; LTI : Limite de tolerância inferior; μ : Média do processo e σ : Desvio padrão estimado.

Na Tabela 1 apresenta-se a classificação do processo através do índice C_{pk} .

Tabela 1. Classificação do processo.

Classificação	Valor de C_{pk}
Capaz	$\geq 1,33$
Razoavelmente capaz	$1 \leq C_{pk} \leq 1,33$
Incapaz	< 1

Fonte: Montgomery (2016).

4 PROCEDIMENTOS METODOLÓGICOS

Nesta seção descreve-se a metodologia utilizada para atingir os objetivos propostos neste trabalho.

4.1 A EMPRESA

A empresa, foco deste estudo, está situada no estado do Paraná, sendo de grande porte e atuação no ramo agroindustrial, em especial na produção de frangos de corte, envolvendo todas as etapas da cadeia produtiva, desde o alojamento até os abates das aves. Possui um frigorífico que realiza o abate diário de aproximadamente 300 mil aves. Atualmente, conta com 6.794 associados e 2.660 funcionários. A sua área de ação está centrada na região Oeste Paranaense, atuando em 12 municípios, sendo eles: Medianeira, Missal, Serranópolis do Iguaçu, São Miguel do Iguaçu, Santa Terzinha de Itaipu, Foz do Iguaçu, Itaipulândia, Santa Helena, Matelândia, Ramilândia, Céu Azul e Vera Cruz do Oeste (Figura 10).

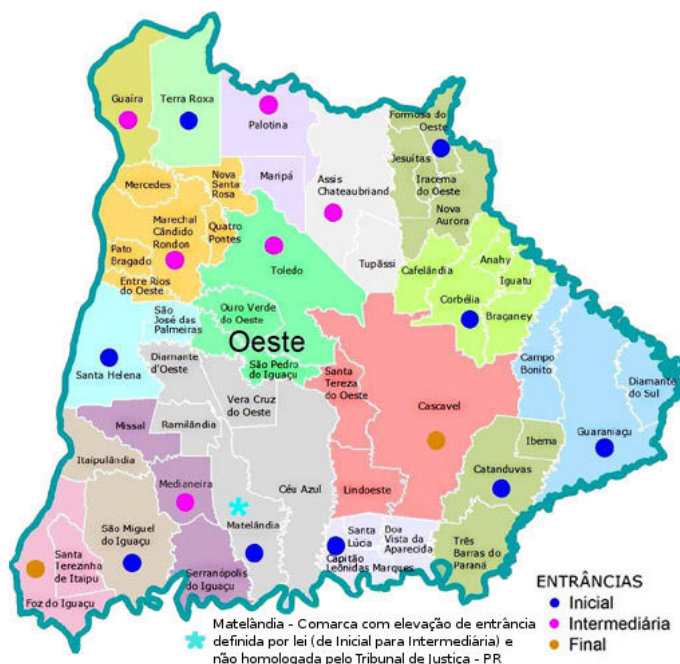


Figura 10. Região Oeste Paranaense

Fonte: Cornelio (2011).

4.2 TIPO DE PESQUISA

A classificação do tipo de pesquisa científica é importante para auxiliar o desenvolvimento de novos métodos científicos, de forma que contribua para a compreensão, análise e discussão de resultados posteriores. Segundo Gerhardt e Silveira (2009) existem diversos tipos de pesquisas, que são classificadas de acordo com a abordagem, natureza, objetivos da pesquisa e procedimentos técnicos.

No que se refere à abordagem a pesquisa é do tipo quantitativa. Segundo Kauark, Manhães e Medeiros (2010) a pesquisa quantitativa reconhece os números e as informações para serem classificadas e analisadas. Requer o uso de técnicas estatísticas (porcentagem, média, mediana, moda, desvio padrão, coeficientes entre outros).

Em relação à natureza, a pesquisa é do tipo aplicada, que conforme GIL (2010) é voltada para a aquisição de conhecimentos para aplicação em uma situação específica.

O projeto de pesquisa, referente à sua classificação, é classificado como pesquisa descritiva, onde é realizada uma descrição de determinada população. Segundo GIL (2010), tem como objetivo identificar relações e associações entre variáveis.

Em relação, ao procedimento técnico adotado, é uma pesquisa operacional, onde são utilizados modelos matemáticos, estatística e algoritmos para ajudar a tomada de decisões (CARDOSO, 2011).

4.3 MINERAÇÃO DE DADOS

A Mineração de Dados geralmente inicia com a limpeza do banco de dados. Gerando um banco de dados mais adequado para Mineração de Dados. Na sequência alguns dados são selecionados em função dos objetivos do trabalho. Posteriormente, os dados passam por um software de Mineração de Dados com seus algoritmos de classificação, predição, regressão, segmentação, associação e padrões sequenciais. Em seguida, tem-se a avaliação, visualização dos resultados e a extração do conhecimento, que são utilizados nas tomadas de decisão (MINUSSI, 2008).

4.4 VARIÁVEIS PRODUTIVAS

Realizou-se, neste trabalho, um levantamento histórico dos dados de dois anos de produção de frangos de corte (01/07/2014 a 31/07/2016) de 919 aviários, nas fases de alojamento e abate. A empresa, ao longo da criação dos frangos, faz o monitoramento de indicadores (Data do Abate, Data do Alojamento, Aves Entregues, Aves Alojadas, Consumo, Idade e Peso total do Lote) que irão compor o Índice de

Eficiência Produtiva (IEP). A cooperativa mede a eficiência produtiva do avicultor por meio do IEP. A remuneração do produtor também está condicionada ao seu desempenho mensurado através do IEP (FIUZA SOBRINHO, 2010). O IEP leva em conta a Conversão Alimentar (CA), o Ganho de Peso Diário (GPD) e a viabilidade (VIAB) do lote. O IEP é obtido por meio da Equação 13:

$$IEP = \frac{\text{Ganho de Peso Diário (GPD)} \times \text{Viabilidade (VIAB)}}{\text{Conversão Alimentar (CA)}} \times 100 \quad (13)$$

Além do IEP, o Peso Médio (PM) é outro índice monitorado pela cooperativa. Este índice é também importante na adequação dos equipamentos automatizados de uma agroindústria de abate de frangos de corte. Portanto, as empresas avícolas buscam uma uniformidade do peso do produto, já que uma baixa uniformidade prejudica o desempenho do processo, a integridade do produto, o comprimento das especificações dos clientes e o aumento de descarte das peças.

O Peso Médio (PM), Equação 14, é a razão entre o peso vivo do lote e o número de frangos entregues (retirados).

$$\text{Peso Médio (PM)} = \frac{\text{Peso total do lote (kg)}}{\text{Aves entregues (AB)}} \quad (14)$$

A Conversão Alimentar (CA), Equação 15, é a relação entre o consumo de ração (kg) e o peso vivo do lote na retirada (kg).

$$\text{Conversão Alimentar (CA)} = \frac{\text{Consumo total de ração (kg)}}{\text{Peso total do lote (kg)}} \quad (15)$$

A Viabilidade (%), Equação 16, é a razão entre o número de frangos retirados e número de frangos recebidos multiplicado por 100.

$$Viabilidade (VIAB) = \frac{Aves Entregues (AB)}{Aves Alojadas (AL)} \times 100 \quad (16)$$

A Idade do lote (Id), Equação 17, é a diferença entre a data do abate e a data do alojamento.

$$Idade (Id) = Data do Abate - Data do Alojamento \quad (17)$$

O Ganho de Peso Diário, Equação 18, é a razão entre o peso médio dos frangos e a idade em dias do abate.

$$Ganho de Peso Diário (GPD) = \frac{Peso Médio (g)}{Idade (Id)} \quad (18)$$

4.5 O SOFTWARE WEKA

O software WEKA, desenvolvido pela Universidade de Waikato na Nova Zelândia, provê vários algoritmos de aprendizado, aplicáveis ao tratamento, treinamento e validação de um conjunto de dados (WITTEN; FRANK, 2005). É escrito na Linguagem Java, dando ao WEKA portabilidade. Este software é distribuído sob a licença *General Public License* (GPL).

O software WEKA é uma das ferramentas de mineração de dados mais utilizada no ambiente acadêmico e se encontra disponível no endereço (<http://www.cs.waikato.ac.nz/ml/weka>).

Entrada de dados:

A entrada de dados do software WEKA (Figura 11) segue o padrão ARFF (*Attribute-Relation File Format*), que é essencialmente um arquivo de texto ASCII delineando instâncias que compartilham atributos. Cada entrada em um conjunto de dados é uma instância da classe *java weka.core.Instance*, e cada uma delas contém atributos específicos, que descrevem um aspecto da instância (WITTEN; FRANK, 2005).

O diagrama mostra um exemplo de arquivo de entrada de dados no formato ARFF. As setas indicam as partes do arquivo:

- Comentário:** As linhas iniciais de texto entre %.
- Nome do Arquivo:** A linha @relation weather.
- Atributos:** As linhas @attribute que descrevem as variáveis.
- Alvo/Variável Classe:** O último valor em cada linha de dados, representando a classe.
- Dados:** O bloco @data contendo as instâncias.

```

% This is a toy example, the UCI weather dataset.
% Any relation to real weather is purely coincidental

@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
  
```

Figura 11 Arquivo de entrada de dados

Fonte: Autoria Própria.

4.6 ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS

Neste trabalho foram avaliados três algoritmos de aprendizado de máquinas: MLP - *Multilayer Perception* (algoritmo baseado em redes neurais), LR - *Linear Regression* (algoritmo baseado em regressão linear) e o *Naive Bayes* (um dos

algoritmos mais eficientes para realizar classificação). Para criar os subconjuntos de dados, de treinamento e teste, foram usados 6000 lotes do banco de dados da cooperativa. Sendo 3000 da linhagem *Coob* e 3000 da linhagem *Coob Slow*. Neste trabalho utilizaram-se dois métodos de fragmentação: o método *Holdhout* onde a base de dados foi dividida com 66% dos dados para treinamento dos algoritmos e 34% para teste e o método *Cross-validation* (Validação cruzada) com 10 partições estratificadas. Nesta técnica os dados são separados em n partições aproximadamente iguais e cada uma delas é utilizada somente uma vez para testes, enquanto as demais são utilizadas para o treinamento (WITTEN; FRANK, 2005). Para validação dos modelos, implementados neste trabalho, utilizou-se um conjunto de 400 lotes de frangos que não participaram das fases de treinamento e teste. Os algoritmos foram treinados no software WEKA 3.8.1.

Número de Neurônios na Camada Oculta:

O número de neurônios da camada oculta (nn) foi definido através da Equação:

$$nn = \frac{nCamadaEntrada + nCamadaSaída}{2} \quad (19)$$

Avaliação dos Resultados:

Os modelos obtidos foram avaliados pelos seguintes parâmetros (WITTEN; FRANK, 2005):

- *Correlation Coefficient* (Coeficiente de Correlação): este coeficiente mede o grau de relacionamento entre duas variáveis. Quanto maior, mais explicativo é o modelo.

- *Mean Absolute Error* (Erro Médio Absoluto) e *Root Mean Squared Error* (Raiz Quadrada do Erro Quadrático Médio): o Erro Médio Absoluto é calculado por meio da diferença entre os valores atuais e os preditos e a Raiz Quadrada do Erro Quadrático Médio é calculada pela média da raiz quadrada da diferença entre o valor calculado e

o valor correto.

- *Relative Absolute Error* (Erro Relativo Absoluto) e *Root Relative Squared Error* (Raiz Quadrada do Erro Quadrático Relativo): Medidas expressas em porcentagem. Menores porcentagens indicam um melhor rendimento dos algoritmos.

Para o modelo de classificação tem-se também:

- *Correctly Classified Instance*: instâncias classificadas corretamente.

Incorrectly Classified Instance: instâncias classificadas incorretamente.

- *Confusion Matrix* (Matriz de Confusão): é a matriz gerada a partir da classificação do conjunto de dados em categorias. Indica a eficiência de um classificador para amostras de diferentes classes. Ela permite visualizar o desempenho de um algoritmo de aprendizado, particularmente em um algoritmo supervisionado, em que cada coluna da matriz representa as instâncias de uma classe prevista, de modo que as linhas representam os casos de um classe real, tornando mais fácil verificar se o algoritmo está confundindo duas classes.

- *Kappa Index*: o índice Kappa (k) (Equação 20) avalia o nível de concordância entre dois conjuntos de dados. Fornece uma idéia do quanto as observações afastam-se das esperadas, indicando o quão legítimas são as interpretações dadas pela Matriz de Confusão.

$$k = \frac{P_0 - P_e}{1 - P_e} \quad (20)$$

Onde: P_0 : taxa de aceitação relativa e P_e : taxa hipotética de aceitação.

Os níveis de concordância são classificados conforme os critérios apresentados na Tabela 2 (Landis e Koch, 1977).

Tabela 2 Níveis de Concordância – Índice Kappa

Valor do Coeficiente Kappa	Nível de Concordância
<0	Não existe concordância
0 - 0,20	Concordância mínima
0,21 - 0,4	Concordância razoável
0,41 - 0,60	Concordância moderada
0,61 - 0,80	Concordância substancial
0,81 – 1	Concordância perfeita

Fonte: Autoria Própria.

Seleção de Atributos:

A seleção de atributos consiste em identificar o subconjunto de atributos mais convenientes para previsão do objeto em estudo. Para isto é necessário eliminar os atributos que sejam pouco significativos, que sejam redundantes ou causem ruídos. Como consequência existe uma melhora no desempenho produtivo e redução do tempo de treinamento e do espaço de armazenamento dos dados.

Existem diversas maneiras de selecionar atributos, mas todas visam a minimização de um conjunto de dados, mantendo os atributos com maior contribuição para o modelo e descartando os demais. Neste trabalho utilizaram-se os algoritmos de seleção CorrelationAttributeEval e CfsSubsetEval (Figura 12) do software WEKA.

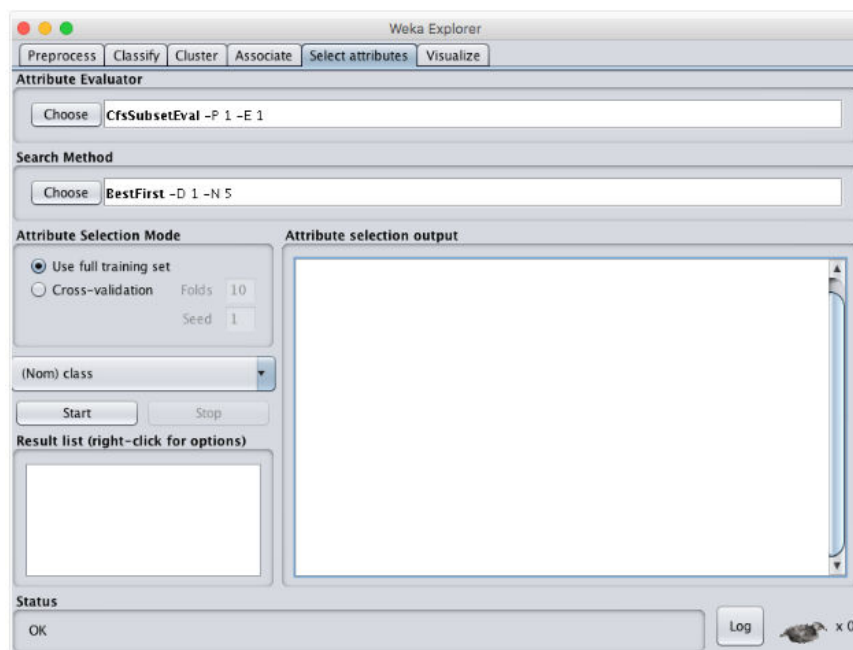


Figura 12 Seleção de atributos – CfsSubsetEval

Fonte: Autoria Própria

A correlação, utilizada utilizada pelos algoritmos de seleção, é uma técnica popular para selecionar atributos relevantes em um banco de dados. Nesta técnica pode-se calcular a correlação entre cada atributo e a variável de saída e utilizar somente os atributos que tem uma correlação positiva ou negativa de moderada a alta (próxima a -1 ou 1) e retirar os atributos com baixa correlação (próxima a zero).

4.7 ÍNDICE DE CAPACIDADE

No Fluxograma apresentado na Figura 13 apresenta-se a metodologia utilizada para o cálculo do índice de capacidade do processo de produção.

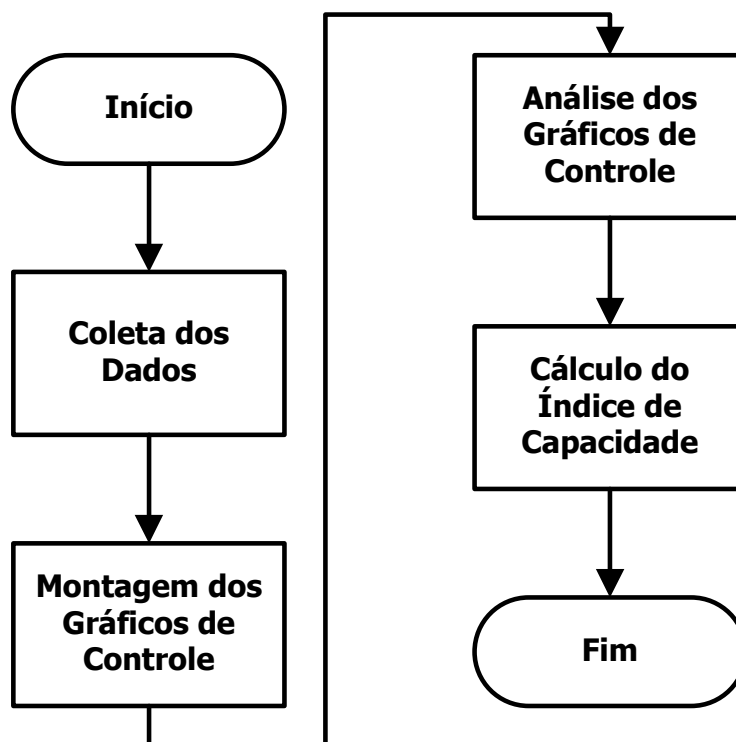


Figura 13 Fluxograma de análise do índice de capacidade do processo de produção de frangos de corte

Fonte: Autoria Própria.

Na sequência apresenta-se cada uma das etapas do fluxograma:

Coleta de Dados – Inicialmente, identificou-se a característica da qualidade (Peso Médio) e separou-se os dados de interesse no banco de dados da empresa.

Montagem dos Gráficos de Controle – Na sequência, definiu-se o tipo de carta de controle a ser usada e realizou-se sua montagem no software.

Análise dos Gráficos de Controle – Em seguida, interpretou-se os resultados obtidos das cartas de controle.

Cálculo do Índice de Capacidade – Finalmente, calculou-se o índice de capacidade C_{pk} do sistema em estudo.

Para a construção das cartas de controle foi utilizada a ferramenta *Quality Control Charts* do software *Statistica® versão 11*.

4.8 PESOS DAS AVES

Os pesos médios das aves de 186 lotes das linhagens Coob e Coob Slow, com uma idade de 45 dias, utilizados neste trabalho para análise de capacidade, foram coletados de forma automática através de uma plataforma de pesagem do modelo 900i da marca Toledo do Brasil. A coleta acontece quando o caminhão carregado chega a agroindústria para realizar a descarga pré-abate das aves.

5 ANÁLISE DE RESULTADOS

5.1 CAPACIDADE DO PROCESSO DE PRODUÇÃO

Foi proposto, neste trabalho, um objetivo de obter informações sobre a capacidade do processo de produção de frangos de corte da empresa alimentícia. Para obter estas informações foram coletados, do banco de dados fornecido pela empresa, 186 lotes de aves das linhagens Coob e Coob Slow com uma idade de 45 dias.

Inicialmente, na análise da estabilidade dos processos, não foram observados pontos acima ou abaixo dos limites superior e inferior de controle, para duas linhagens. Porém, apresentaram outros padrões de causas especiais (Figuras 14 e 15).

		PM ; Runs Tests (CS)	
		X Chart	
		Center line: 2,814833 Sigma: 0,089797	
Zones A/B/C: 3,000/2,000/1,000 * Sigma		from sample	to sample
Tests for special causes (runs rules)			
9 samples on same side of center		OK	OK
6 samples in row in/decreasing		OK	OK
14 samples alternating up & down		OK	OK
2 of 3 samples in Zone A or beyond		1	3
		44	46
		69	71
4 of 5 samples in Zone B or beyond		44	48
		67	71
		75	79
15 samples in Zone C		OK	OK
8 samples beyond Zone C		75	82

Figura 14 Outros padrões de causas especiais - Coob Slow

Fonte: Autoria própria.

PM ; Runs Tests (CB)	
X Chart	
Center line: 2,795119 Sigma: 0,094941	
Zones A/B/C: 3,000/2,000/1,000 * Sigma	
Tests for special causes (runs rules)	
9 samples on same side of center	from sample 33 to sample 41
6 samples in row in/decreasing	OK OK
14 samples alternating up & down	OK OK
2 of 3 samples in Zone A or beyond	87 89
	92 94
4 of 5 samples in Zone B or beyond	24 28
	50 54
15 samples in Zone C	OK OK
8 samples beyond Zone C	OK OK

Figura 15 Outros padrões de causas especiais – Coob

Fonte: Autoria própria.

Como este é um estudo inicial do processo, recalcularam-se os limites de controle excluindo as causas especiais, que numa implantação normal do CEP, teriam sido corrigidas (Figuras 16 e 17).

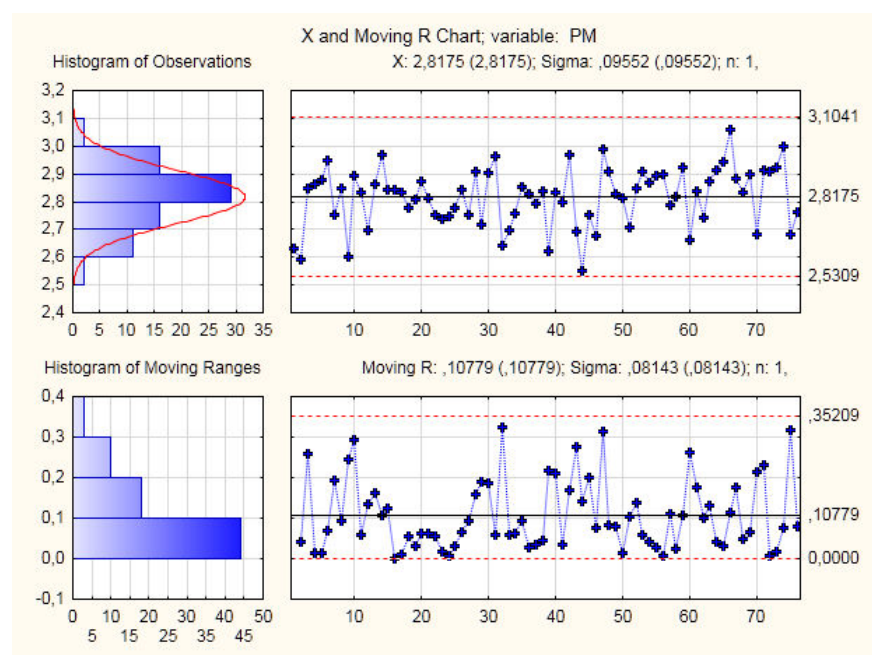


Figura 16 Gráfico Xi - MR – Coob Slow – Peso Médio

Fonte: Autoria própria.

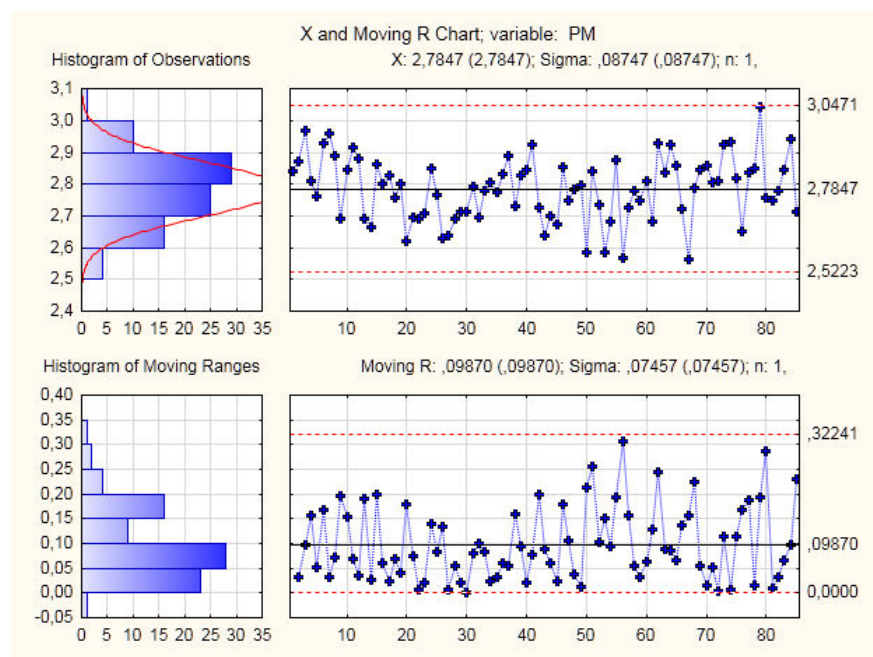


Figura 17 Gráfico Xi - MR – Coob – Peso Médio

Fonte: Autoria própria.

Quando se trata do estudo de capacidade de um processo, tem que se ter cuidado especial com a normalidade dos dados. Os dados das duas linhagens (peso médio) foram submetidos ao teste de *Shapiro-Wilk* para verificação de sua normalidade. Como os valores de p foram maiores que 0.05 ($p > 0.05$), as distribuições dos dados foram consideradas normais (Figuras 18 e 19).

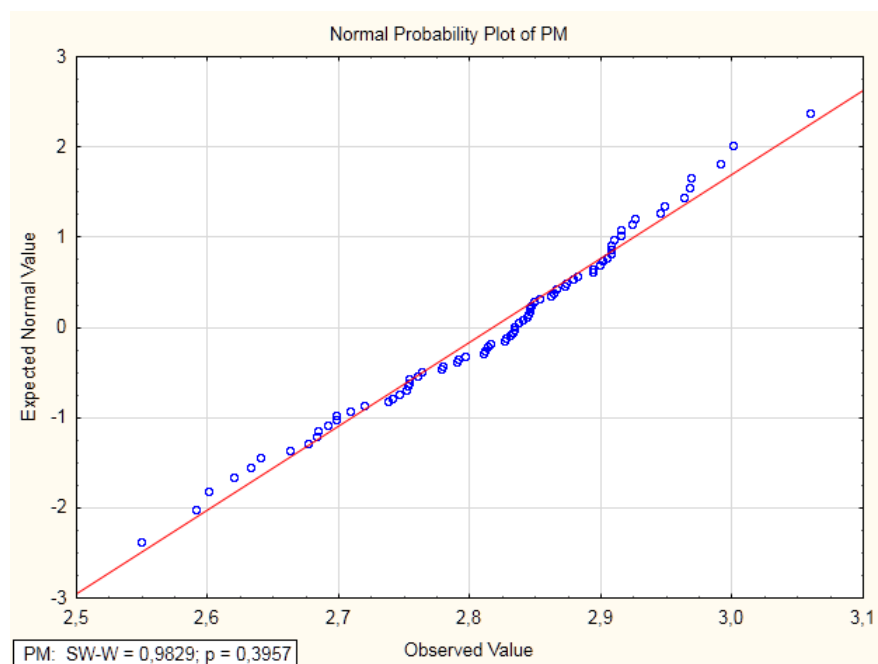


Figura 18 Gráfico de Probabilidade Normal – Coob Slow

Fonte: Autoria própria.

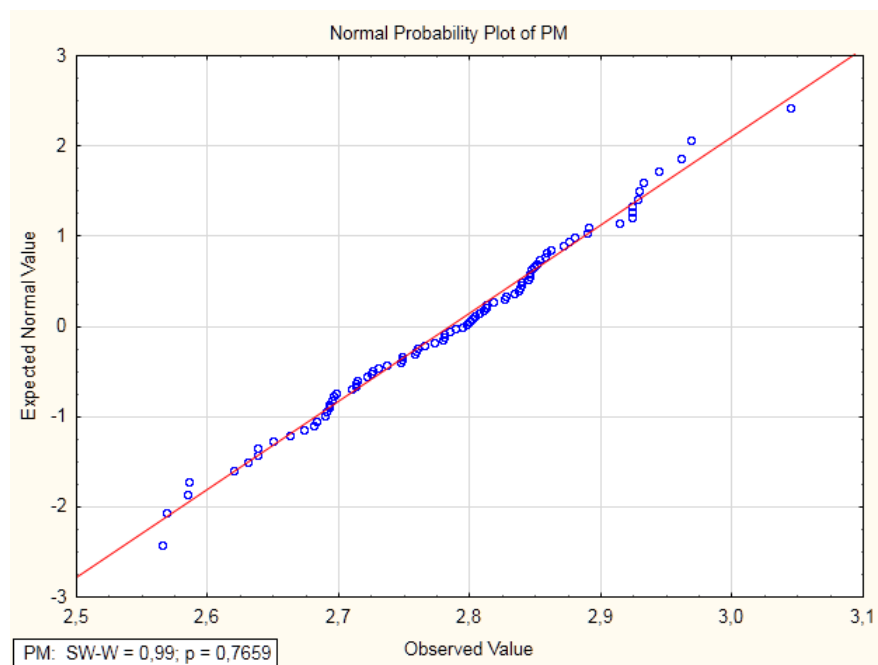


Figura 19. Gráfico de Probabilidade Normal – Coob

Fonte: Autoria própria.

Verificada a normalidade dos processos quantificaram-se suas capacidades

usando o índice C_{pk} . A empresa, segundo o Departamento de Engenharia, tem um valor alvo, para os pesos médios dos frangos, de 2,75 kg com uma variação de 0,15 kg. Nas Figuras 20 e 21 apresentam-se os histogramas e os índices de capacidade.

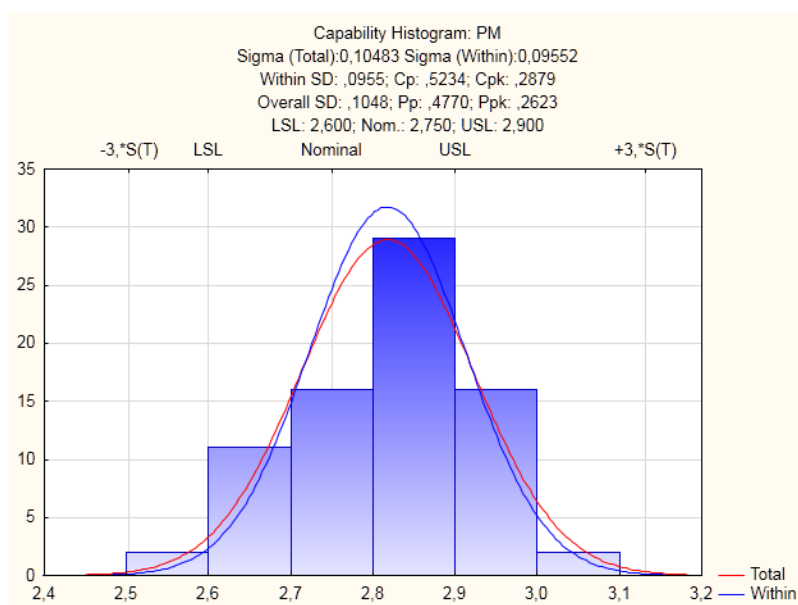


Figura 20 Histograma e Índice de Capacidade C_{pk} — Coob Slow

Fonte: Autoria própria.

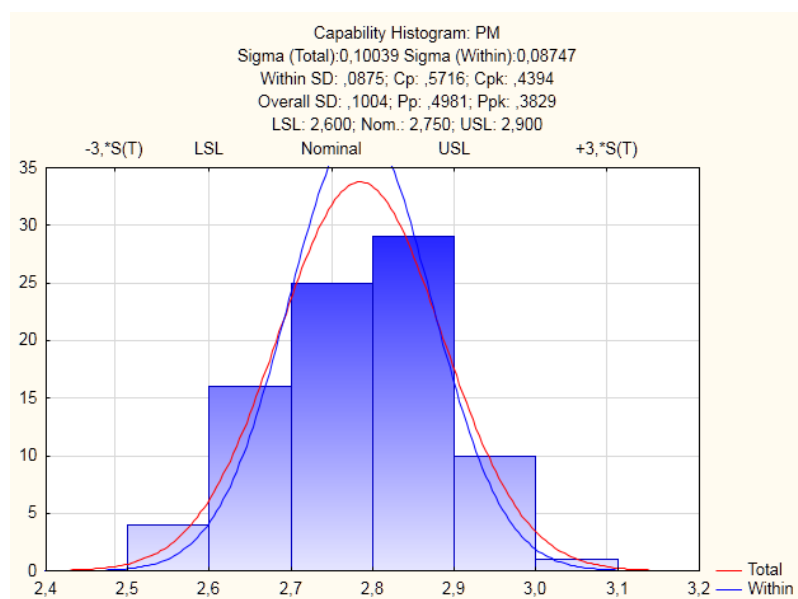


Figura 21 Histograma e Índice de Capacidade C_{pk} — Coob

Fonte: Autoria própria.

Nas Figuras 20 e 21, é possível observar que os processos não são capazes, segundo o cálculo do C_{pk} , pois seus valores calculados ficaram bem abaixo de 1,33. Observa-se, também, que os processos possuem médias deslocadas para acima dos alvos.

5.2 MODELOS DE PREVISÃO

Seleção de Atributos:

Para encontrar os atributos relevantes para as variáveis de resposta utilizou-se o módulo de seleção de atributos do WEKA. Nesse ambiente foi utilizado o algoritmo *CorrelationAttributeEval* com o modo de busca *Ranker*. O modo *Ranker* faz uma avaliação individual dos atributos e os organiza com sua ordem de importância.

Na Tabela 3 apresentam-se os resultados obtidos, classificação dos atributos, para as variáveis de resposta: Índice de Eficiência Produtiva e Peso Médio.

Tabela 3 Classificação dos atributos

Filtro de classificação por correlação			
PM		IEP	
Classificação	Atributos	Classificação	Atributos
0,89288	CONS	0,864	GPD
0,83283	GPD	0,3487	VIAB
0,64863	ID	0,2088	CONS
0,05182	LIN	0,0769	AB
0,00421	AB	0,0604	AL
0,00263	AL	0,028	LIN
-0,01521	CA	-0,118	ID
		-0,7246	CA

Fonte: Autoria própria.

As variáveis de entrada relevantes para cada uma das variáveis de resposta são apresentadas na Tabela 4.

Tabela 4 Variáveis de entrada e saída para os algoritmos

Saída	Entradas				
Peso Médio	CONS	GPD	ID	-	-
Índice de Eficiência Produtiva	GPD	VIAB	CONS	ID	CA

Fonte: Autoria própria.

5.3 ALGORITMOS

Utilizando o software WEKA, implementou-se duas redes neurais seguindo o *modelo Multilayer Perceptron* com algoritmo de aprendizado por retropropagação de erros (MLP). As redes modeladas foram especificadas como: MLP 5-3-1 com 5 neurônios na camada de entrada, 3 neurônios na camada oculta e 1 na camada de saída e MLP 3-2-1 com 3 neurônios na camada de entrada, 2 neurônios na camada oculta e 1 na camada de saída (Figura 22). Todos os neurônios, das redes neurais implementadas neste trabalho, tem como função de ativação a sigmóide.

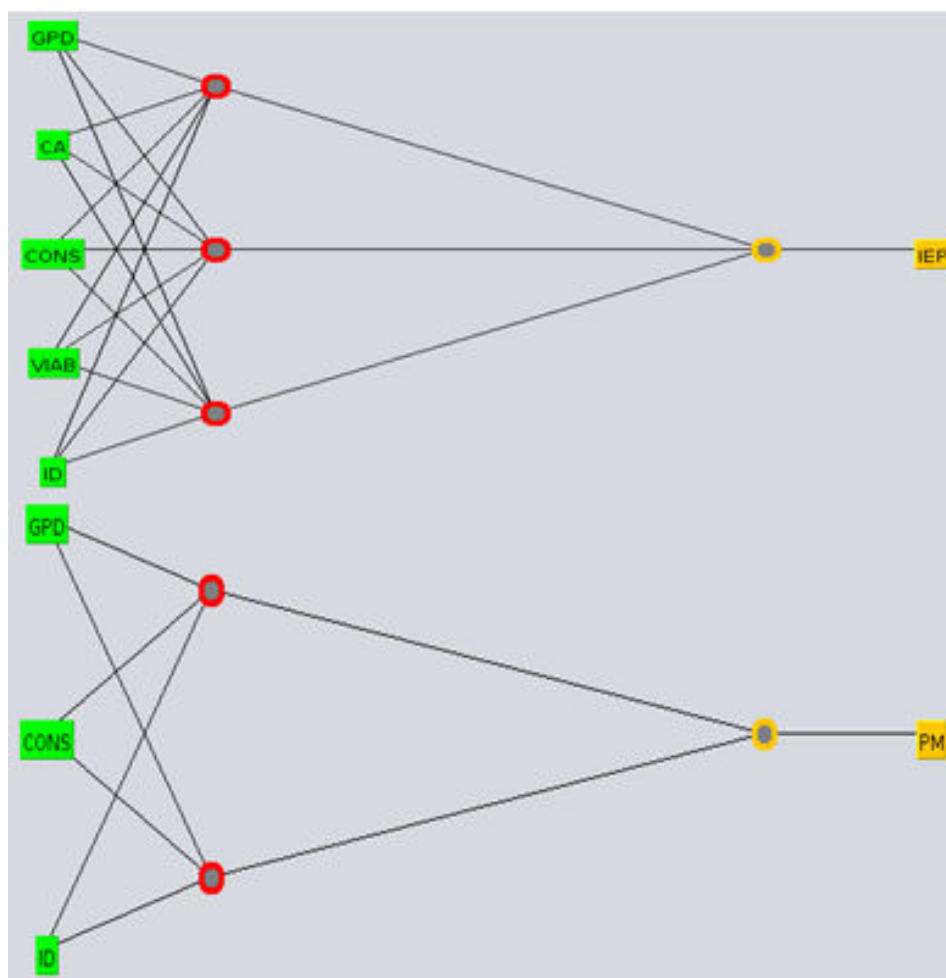


Figura 22 Redes MLP 5-3-1 e Rede MLP 3-2-1

Fonte: Autoria própria.

As equações de regressão obtidas, para os dois modelos, por meio do algoritmo *Linear Regression* são apresentadas nas Equações 21 e 22.

$$PM = 0,0453GPD + 0,0046CONS + 0,0611ID + 2,7877 \quad (21)$$

$$IEP = 9,6214GPD - 79,7068CA - 49,9281CONS + 3,7623VIAB + 5,0565ID - 462,6132 \quad (22)$$

A Tabela 5 apresenta os resultados obtidos, dos dois algoritmos, para as variáveis de resposta Índice de Eficiência Produtiva (IEP) e Peso Médio (PM).

Tabela 5. Resultados da aplicação dos algoritmos MLP e LR

	IEP		PM	
	LR	MLP	LR	MLP
Coeficiente de Correlação	0,9995	0,9994	0,9996	0,9995
Erro médio absoluto	0,5334	0,6531	0,0035	0,0034
Erro médio quadrático	0,7181	0,7918	0,0045	0,004
Erro absoluto relativo (%)	2,941	3,6006	2,8837	2,7622
Raiz quadrada do erro médio quadrático (%)	3,1635	3,488	2,994	2,8079

Fonte: Autoria própria.

Verifica-se, por meio dos resultados apresentados na Tabela 5, que os dois algoritmos apresentam praticamente os mesmos resultados de coeficiente de correlação. A análise dos erros permitem verificar que os melhores resultados foram obtidos usando o algoritmo de previsão LR para a variável Índice de Eficiência Produtiva (IEP) e o algoritmo MLP para a variável Peso Médio (PM).

Dados Preditos versus Dados Observados:

Na sequência, avaliou-se a dispersão dos valores preditos pelos dois algoritmos em relação aos valores observados para as duas variáveis de resposta (Figura 23).

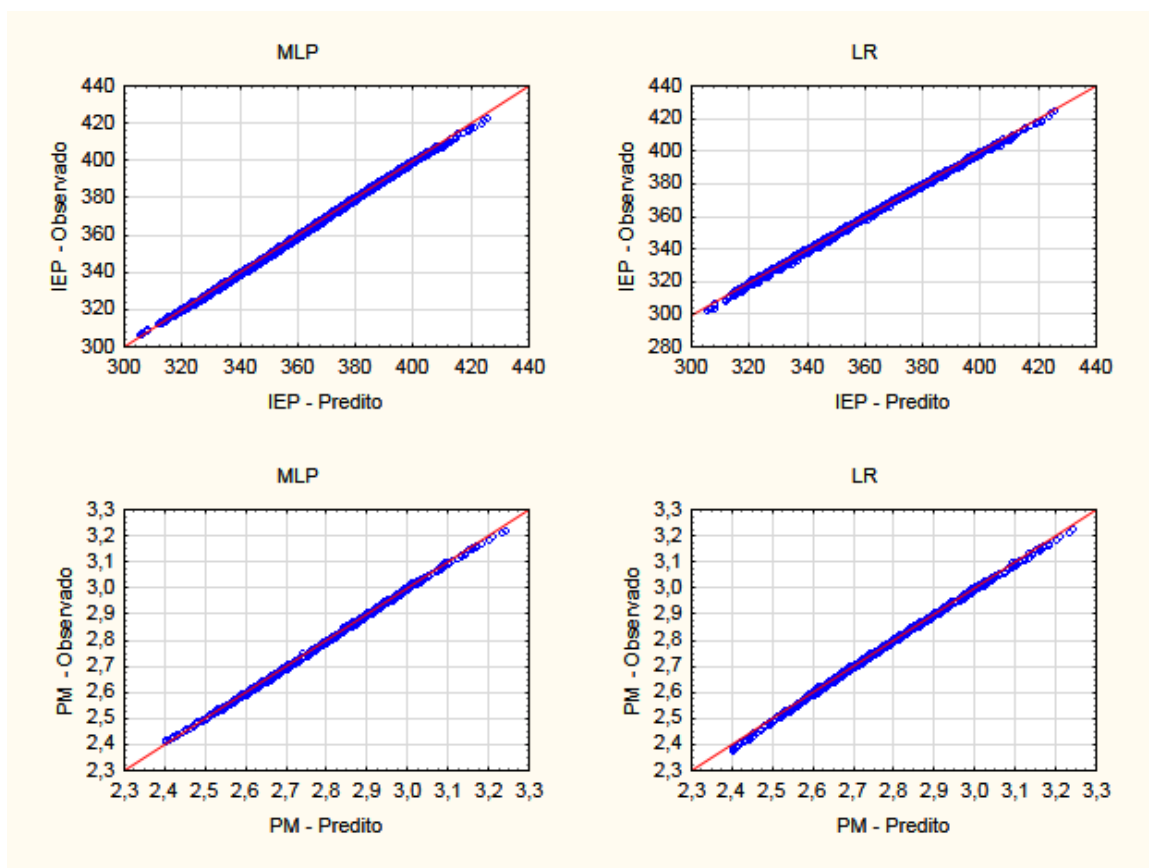


Figura 23. Dados preditos em função de dados observados

Fonte: Autoria própria.

Observa-se que os dados experimentais apresentam uma boa correlação com os dados preditos para as duas variáveis produtivas, já que os pontos estão próximos a linha de ajuste.

Teste dos Modelos:

Para testar os modelos compararam-se resultados obtidos dos algoritmos com resultados obtidos de 400 lotes de frango (Figura 24).

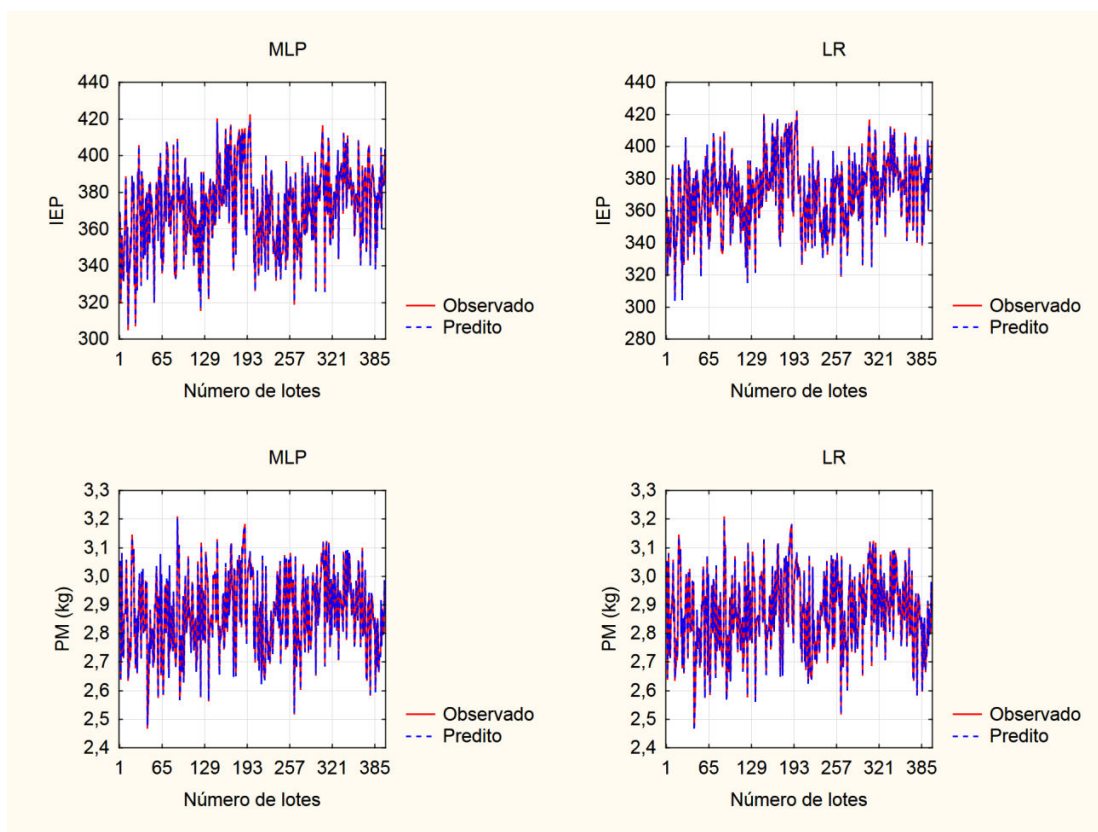


Figura 24 Dados preditos em função de dados observados – 400 lotes

Fonte: Autoria própria.

Pode-se observar, por meio dos gráficos, apresentados na Figura 24, que existe uma boa aproximação entre os resultados reais e os previstos pelos dois algoritmos.

É relevante observar que o método *Cross-validation* apresentou melhores resultados, em termos de erro, que o método *Holdhout* para a variável IEP (Tabela 6).

Tabela 6. Resultados da aplicação do algoritmo MLP – *Cross-validation e Holdhout*

	IEP	
	MLP - <i>Cross-validation</i>	MLP - <i>Holdhout</i>
Coeficiente de Correlação	0,9996	0,9994
Erro médio absoluto	0,4558	0,6531
Erro médio quadrático	0,6132	0,7918
Erro absoluto relativo (%)	2,5187	3,6006
Raiz quadrada do erro médio quadrático (%)	2,7397	3,488

Fonte: Autoria própria.

Na Figura 25 apresenta-se os resultados obtidos com os 400 lotes utilizando o método *Cross-validation*.

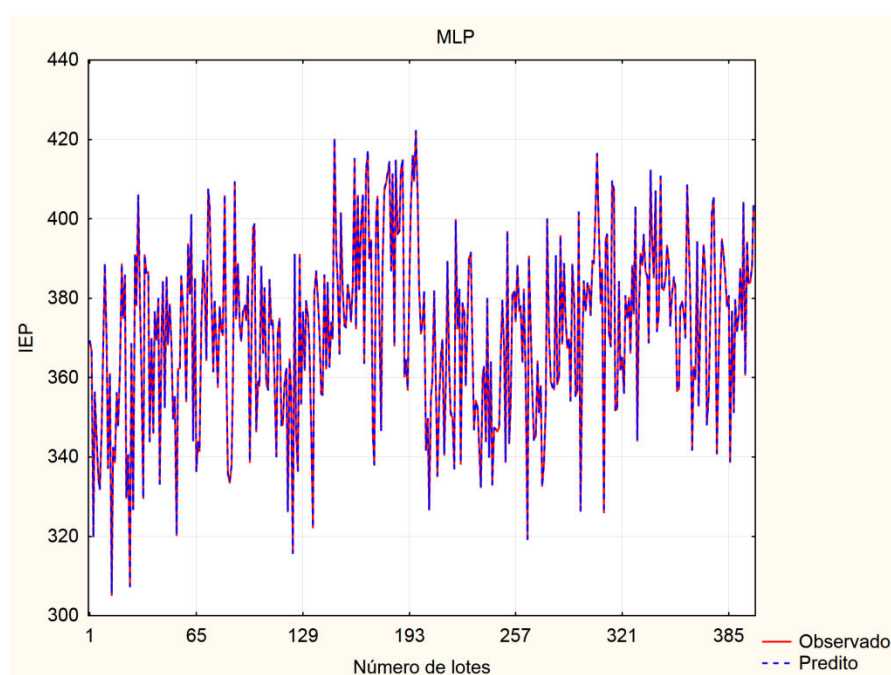


Figura 25 Dados preditos em função de dados observados – MLP e *Cross-validation*

Fonte: Autoria própria.

Realí (2004) também obteve previsões muito boas, aplicando um modelo de redes neurais, para o Índice de Eficiência Produtiva (IEP) e para o Peso Médio (PM)

de 1041 lotes para os anos de 2001 e 2002, em uma Intervenção Agrícola do Rio Grande do Sul, utilizando o Software NeuroShellPredictor. Obtendo, em suas análises, um coeficiente de correlação de 0,999951 para o Peso Médio e um coeficiente de 0,996006 para o Índice de Eficiência Produtiva.

5.4 MODELO DE CLASSIFICAÇÃO

Para encontrar os atributos relevantes para variável de resposta linhagem (LIN) utilizou-se o módulo de seleção de atributos do WEKA. Nesse ambiente foi utilizado o algoritmo *CfsSubsetEval* com o modo de busca *BestFirst*. Este algoritmo avalia o valor de um subconjunto de atributos considerando a habilidade preditiva individual de cada característica juntamente com o grau de redundância entre eles. O modo *BestFirst* é o modo de pesquisa padrão do WEKA. Os atributos selecionados, para a classe LIN, foram a Conversão Alimentar (CA) e o Peso Médio (PM).

Ao rodar o classificador Naive Bayes, por meio do método *cross-validation*, o software oferece um relatório de desempenho (Tabela 5.5), onde gera o índice kappa, que no caso do treinamento/validação foi de 0,957. Ou seja, uma acurácia muito boa. Na matriz de confusão, a classe CB teve 127 amostras erroneamente classificada como CS e para classe CS 2 amostras como CB. Já para o conjunto de teste obteve-se um índice kappa de 0,985 e somente 3 amostra CB erroneamente classificada como CS.

Tabela 7 Resultados da aplicação do algoritmo Naive Bayes

Sumário	Treinamento/Validação	Teste
Instâncias corretamente classificadas	5871 – 97,85%	397 - 99,25%
Instâncias incorretamente classificadas	129 – 2,15%%	3 - 0,75%
Estatística Kappa	0,957	0,985
Erro médio absoluto	0,0495	0,025
Erro médio quadrático	0,1448	0,0958
Erro absoluto relativo (%)	9,8927	5,0062
Raiz quadrada do erro médio quadrático (%)	28,9604	19,1662
Total de Instâncias	6000	400
Matriz de confusão	$\begin{matrix} & a & b \\ a & 2873 & 127 \\ b & 2 & 2998 \end{matrix}$ $\begin{matrix} a = CB \\ b = CS \end{matrix}$	$\begin{matrix} & a & b \\ a & 197 & 3 \\ b & 0 & 200 \end{matrix}$ $\begin{matrix} a = CB \\ b = CS \end{matrix}$

Fonte: Autoria própria.

6 CONCLUSÃO

Este trabalho teve por objetivo aplicar as técnicas de Controle Estatístico de Processo (CEP) e algoritmos de mineração de dados na análise, previsão e classificação de variáveis produtivas que compõem as características das aves a serem abatidas em uma cooperativa localizada na região oeste paranaense.

A empresa forneceu um banco de dados com um histórico de movimentação de 2 anos, contendo dados desde o alojamento até a fase de abate. Neste estudo, utilizaram-se as seguintes variáveis: Ganho de Peso Diário, Consumo, Idade, Conversão Alimentar, Aves Alocadas, Aves Abatidas, Linhagem, Peso Médio e Índice de Eficiência Produtiva.

Inicialmente, aplicou-se técnicas de Controle Estatístico de Processo na análise do peso médio de 186 lotes de aves das linhagens Coob e Coob Slow, com uma idade de 45 dias. Na aplicação das cartas de controle observou-se que o processo de produção de frango apresentava causas especiais. Observou-se, com a retirada das causas especiais, que o processo estava sob controle estatístico e os dados seguiam uma distribuição normal, condição necessária para o estudo da capacidade do processo.

Pela análise dos resultados dos índices de capacidade percebeu-se que os processos produtivos, para o caso analisado, não são precisos para produzir frangos com as qualidades requeridas pela empresa. Portanto, os processos de produção, das duas linhagens dos aviários estudados, precisam ser revisados para levar a média para o alvo.

Com as variáveis produtivas, obtidas do banco de dados da empresa, construíram-se modelos de previsão baseados em dois algoritmos de análise de dados: Redes Neurais Artificiais (MLP) e Regressão Linear (LR). Os modelos obtidos, por meio dos algoritmos, estão bem ajustados como pode ser observado pela correlação entre os dados preditos e observados.

Observou-se nesse trabalho que o consumo e o ganho de peso diário foram

os parâmetros produtivos que mais contribuíram para as variáveis de resposta (Peso Médio e Índice de Eficiência Produtiva). Observou-se, também, que a Linhagem das aves não teve praticamente influência sobre as variáveis de resposta.

Quando os modelos, implementados para prever as variáveis Peso Médio e Índice de Eficiência Produtiva, foram aplicados a 400 lotes de frangos, as previsões foram bem precisas e as diferenças entre valores reais e previstos foram pequenas. Portanto, a proximidade entre valores preditos e reais demonstram a boa capacidade de generalização dos modelos implementados neste trabalho.

Observou-se que o algoritmo de *Naive Bayes*, para o conjunto de validação, teve 99,25% de instâncias corretamente classificadas. Apresentando, portanto, um bom desempenho na classificação das linhagens das aves.

É importante ressaltar que modelos gerados só têm utilidade para a empresa onde foi realizado o estudo.

REFERÊNCIAS

ALBINO, L.F.T.; TAVERNARI, F.C. **Produção e manejo de frangos de corte**. Viçosa: UFV, 2008.

ALENCAR, J.; LOPES, C. E.; SOUZA, M. **Monitoramento do processo de compressão de comprimidos de captopril utilizando controle estatístico de processo**. Rev. Bras. Farm, 2007.

Associação Brasileira de Proteína Animal (ABPA). **Relatório Anual Completo**. Disponível em:< http://abpa-br.com.br/files/RelatorioAnual_UBABEF_2015_DIGITAL.pdf>. Acesso em: 12 outubro 2016, 20:41

AZARIAS P.; MATOS, S. N. & SCANDELARI, L.. **Aplicação da mineração de dados para a Geração do Conhecimento: um experimento prático**. In: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 5, 2009, Niterói. Anais... Rio de Janeiro: USP, 2009. p. 102-119.

BARCZSZ, S.S.; LIMA FILHO, D.O. **Agroindústria exportadora de frango de corte Sul-Mato-Grossense e os aspectos de internacionalização**. Revista em Agronegócios e Meio Ambiente, v.2, n.2, p. 9-33, mai./ago. 2009

BELUSSO, Diane; HESPANHOL, Antonio Nivaldo. **A evolução da avicultura industrial brasileira e seus efeitos territoriais**. Maringá-PR, Brasil. Revista Percurso, v. 2, n. 1, 2177-3300, 2010.

BERTHOLD, M. ; Hand, D. J. **Intelligent Data Analysis - Ana Introduction**. 2. Ed. Springer-Verlag. 2010.

BINOTI, M. L. M. S. **Rede neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto**. [Dissertação]. Viçosa: Universidade Federal de Viçosa; 2010. 54 p.

CALDARELLI, C.E.; CAMARA, M.R.G. **Efeitos das variações cambiais sobre os preços da carne de frango no Brasil entre 2008 e 2012**. Revista Economia e Sociologia Rural, v. 51, n. 3, 575-590, jul/set, 2013

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf>. Acesso em: 12 outubro 2016, 21:35.

CARDOSO, Andréa. **Fundamentos da Pesquisa Operacional**. Universidade Federal de Alfenas (Unifal), 2011. Disponível em: . Acesso em: 01 maio 2017

CORDEIRO, A. F. S; BARACHO, M. S; NASS, I. A; NASCIMENTO, G. R. **Uso de mineração de dados para identificação de fatores que influenciam o grau de lesões de perna em frangos de corte** . 2012, vol.32, n.4, pp.642-649. ISSN 0100-6916. Disponível em: < http://www.scielo.br/scielo.php?pid=S0100-69162012000400003&script=sci_abstract&lng=pt>. Acesso em: 10 maio 2018, 20:47.

CORNÉLIO, Sóstenes Carvalho. **Sustentabilidade no setor de construção civil da Região Oeste do Paraná** . 2011. Disponível em: < <https://acervodigital.ufpr.br/handle/1884/26257>>. Acesso em: 10 maio 2017, 14:33.

COSTA, Antonio Fernando Branco; EPPRECHT, Eugenio Kahn; CARPINETTI, Luiz Cesar Ribeiro. **Controle estatístico de qualidade**. 2. ed. São Paulo: Atlas, 334 p., 2010.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária: **Linhagens, 2017.**

Disponível em: <

https://www.agencia.cnptia.embrapa.br/gestor/frango_de_corte/arvore/CONT000fc68q25k02wx5eo0a2ndxy8lyc3jr.html> Acesso em: 01/05/2017.

ESPÍNDOLA, C.J. **Trajetórias do progresso técnico na cadeia produtiva de carne de frango do Brasil.** Revista Geosul, v. 27, n. 53, p. 89-113, jan./jul., 2012.

FAYYAD UM, Shapiro GP.; Smyth P.; Uthurusamy R. **Advances in knowledge discovery and data mining.** Menlo Park, Calif.: AAAI Press: MIT Press; 1996; 611p.

FAYYAD, U; PIATETSKY-SHAPIRO, G; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases.** American Association for Artificial Intelligence, 1996.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo. **Métodos de pesquisa.** Porto Alegre: Editora da Ufrgs, 2009. 120 p.

GIL, Antonio Carlos, 1946 – **Como elaborar projetos de pesquisa** – 5 ed., São Paulo, Atlas, 2010. 184p.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. **Data mining: um guia prático, conceitos, técnicas, ferramentas, orientações e aplicações.** 2 ed., São Paulo: Elsevier; 2015.

HAN, J; KAMBER, M. **Data Mining: Concepts and Techniques.** 3. Ed. Elsevier, 2011.

HAND, D; MANNILA, H; SMYTH, P. **Principles of Data Mining.** MIT Press, 2001.

HAYKIN, S. **Redes Neurais, Princípios e prática**. 2. ed. [S.l.]: Bookman, 2003

KAUARK, Fabiana; MANHÃES, Fernanda Castro; MEDEIROS, Carlos Henrique. **Metodologia da pesquisa**. Itabuna, Via Litterarum: Via Litterarum Editora, 2010. 88 p.

LANDIS, J.R.; KOCH, G.G. **The measurement of observer agreement for categorical data**. *Biometrics*, v.33, n.1, p. 159-174, 1977.

LIMA, C.E; MARTINS, T. C.; SOLDADA, G. V.; SILVA, R. S. **Caracterização das exportações e da competitividade internacional do complexo de carnes brasileiras**. In: ENCONTRO DE ECONOMIA CATARINENSE INOVAÇÃO E DESENVOLVIMENTO, 6, 2012, Joinville. Anais.Joinville: APEC, 2012.

LIMA, M. G. F; RODRIGUES, L. H. A. **Árvore de decisão aplicada em dados de incubação de matrizes de postura Hy-Line W36**. 2010, vol.34, n.6, pp.1550-1556. ISSN 1413-7054. Disponível em: < http://www.scielo.br/scielo.php?pid=S1413-70542010000600028&script=sci_abstract&lng=pt>. Acesso em: 06 maio 2018, 15:54.

MATOS, G.; Chalmeta R., Coltell O. **Metodología para la extracción del conocimiento empresarial a partir de los datos**. *Inf Tecnol*. 2006;17(2):81-8.

MCCUE, C. **Data Mining and Predictive Analysis - Intelligence Gathering and Crime Analysis**. 2. ed. Elsevier, 2015.

MINUSSI, J. A.. **New Corporate Credit Scoring Models in the Context of the Basel II Accord: A Brazilian Case Study**. Tese de Doutorado do Departamento de Ciências Administrativas da Universidade de Lancaster, 361 páginas, 2008.

MONTGOMERY, D. C. **Introdução ao controle estatístico de qualidade**. 7. Ed. Rio de Janeiro: LTC, 2016.

G. R. R. NASCIMENTO; I. A. NAAS; D. F. PEREIRA; W. M. DUTRA JUNIOR, A. P. A. MAIA **Previsão de Conforto Térmico de Frangos de Corte Utilizando Mineração de Dados** Revista Brasileira de Engenharia de Biosistemas 2011.

OLIVEIRA, D.R.M.S.; NÄÄS, I.A. **Issues of sustainability on the Brazilian broiler meat production chain**. In: INTERNATIONAL CONFERENCE ADVANCES IN PRODUCTION MANAGEMENT SYSTEMS, 2012, Rhodes. Anais. Competitive Manufacturing for Innovative Products and Services: proceedings, Greece: Internacional Federation for Information Processing, 2012.

RAMOS, A. W. **Mantendo o processo sob controle**. In: ROTONDARO, R. G. **Seis Sigma: Estratégia gerencial para a melhoria de processos, produtos e serviços**. São Paulo: Atlas, 2013.

REALI, E. H. **Utilização de inteligência artificial - (Redes Neurais Artificiais) no gerenciamento da produção de frangos de corte**. Dissertação de Mestrado - UFRS, RS, 2004.

RIBEIRO, J. L. D.; CATEN, C. S. T. **Controle Estatístico do Processo**. 2012. 172 f. Monografia (Pós Graduação em Engenharia) – Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2012.

RIPPEL, Ricardo. **Migração e desenvolvimento econômico no Oeste do estado do Paraná: uma análise de 1950 a 2000**. Campinas-SP, Brasil. 2005. Disponível em: <
http://repositorio.unicamp.br/bitstream/REPOSIP/280464/1/Rippel_Ricardo_D.pdf>.
Acesso em: 08 julho 2018, 21:27.

RODRIGUES, V. C; VIEIRA, F. M. C; SILVA, I. J O. **Mineração de dados para estimativas de mortalidade pré-abate de frangos de corte**. Piracicaba-SP, Brasil. 2010. ISBN 18854494. Disponível em: <
http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0004-05922013000300015>. Acesso em: 08 maio 2018, 19:19.

SILVA, R. A. **Estudo preliminar da aplicação de redes neurais no sistema de controle de temperatura em uma planta didática**. [Trabalho de Conclusão de Curso]. Niteroi: Universidade Federal Fluminense; 2015.

FIUZA SOBRINHO, R. **Competitividade na cadeia de valor da Avicultura de corte**. Dissertação de Mestrado - UNIOESTE, PR, 2010.

UBA, União Brasileira de Avicultura. **Protocolo de Boas Práticas de Produção de Frangos**. São Paulo-SP, 2008.

VOILÀ, M.; TRICHES, D. **A cadeia de carne de frango: uma análise dos mercados brasileiro e mundial de 2002 a 2010**. Instituto de Pesquisas Econômicas e Sociais, n.44, 2013.

ZHOU, Z.-H. **Three perspectives of data mining**. Artificial Intelligence Journal, p. 139–146, 2003.

WITTEN, I. H.; FRANK, E. **Data Mining: practical machine learning tools and techniques**. São Francisco: Elsevier, 2005.