



**UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**



OTON CRISPIM BRAGA

**CREVETTIC: UMA PLATAFORMA DE *BUSINESS*
INTELLIGENCE PARA GESTÃO DE CARCINICULTURA**

MOSSORÓ

2020

OTON CRISPIM BRAGA

**CREVETTIC: UMA PLATAFORMA DE *BUSINESS*
INTELLIGENCE PARA GESTÃO DE CARCINICULTURA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação - associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Francisco Milton Mendes Filho - UFERSA

Coorientador: Prof. Dr. Antônio Mauro Barbosa de Oliveira - IFCE

MOSSORÓ

2020

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

B813c Braga, Oton Crispim.
CREVETTIC: UMA PLATAFORMA DE BUSINESS
INTELLIGENCE PARA GESTÃO DE CARCINICULTURA / Oton
Crispim Braga. - 2020.
66 f. : il.

Orientador: Francisco Milton Mendes Neto.
Coorientador: Antônio Mauro Barbosa de
Oliveira.
Dissertação (Mestrado) - Universidade Federal
Rural do Semi-árido, Programa de Pós-graduação em
, 2020.

1. Carcinicultura. 2. Sistema de Apoio a
Decisão. 3. Business Intelligence. 4. Análise
de Dados. I. Mendes Neto, Francisco Milton ,
orient. II. Oliveira, Antônio Mauro Barbosa de ,
co-orient. III. Título.

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

OTON CRISPIM BRAGA

CREVETTIC: UMA PLATAFORMA DE *BUSINESS INTELLIGENCE* PARA GESTÃO DE CARCINICULTURA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.

APROVADA EM: 18 / 10 / 2020.

BANCA EXAMINADORA



Prof. Dr. Francisco Milton Mendes Filho – UFERSA
Presidente



Prof. Dr. Antônio Mauro Barbosa de Oliveira – IFCE
Primeiro Membro



Prof. Dr. Lenardo Chaves e Silva – UFERSA
Segundo Membro



Prof. Dra. Carina Teixeira de Oliveira – IFCE
Terceiro Membro

Dedico essa produção intelectual e científica à minha família, amigos, colegas e professores, que me apoiaram, tecnicamente e emocionalmente. Sem eles eu não teria capacidade de concluir.

AGRADECIMENTOS

Agradeço aos meus orientadores e amigos, professor Milton Mendes pelo esforço e disposição durante o desenvolvimento deste trabalho, e professor Mauro Oliveira pelos seus conselhos e incentivos.

Agradeço também à família LES-UFERSA (Laboratório de Engenharia de Software da UFRSA), onde encontrei apoio, amigos, conhecimento e infraestrutura para desenvolver este trabalho.

Agradeço a equipe administrativa e professores do PPgCC (Programa de Pós-Graduação em Ciência da Computação UFRSA/UERN), que contribuíram com a minha formação intelectual e formação de carácter.

Agradeço a (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) CAPES por ter me proporcionado apoio financeiro para o desenvolvimento desse trabalho, de vital importância para minha manutenção no programa.

Agradeço também a todos os demais que colaboraram direta e indiretamente para realização desse trabalho, incluindo minha família e amigos. Reconheço que muita coisa acontece nos bastidores.

EPÍGRAFE

“A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original.”

(Albert Einstein)

RESUMO

A produção de camarão em cativeiro é um importante setor na economia do Nordeste do Brasil, principalmente nos estados do Ceará e Rio Grande do Norte, que juntos produzem 66% da produção nacional anual, cerca de 27,2 mil toneladas no total. No entanto, o aparecimento de novas doenças tem obrigado produtores a realizarem um acompanhamento mais efetivo em relação ao manejo e à saúde dos camarões. São coletadas diversas informações a fim de identificar previamente sinais que possam causar prejuízos de alguma natureza à produção, como desenvolvimento lento e aumento da taxa de mortalidade. Contudo, muitas fazendas ainda realizam o controle dessas informações de forma ineficiente, usando papel ou por meio de planilhas que acabam crescendo consideravelmente, dificultando a análise dos dados por seres humanos. Desse modo, a tomada de decisão nas fazendas acaba sendo realizada com base na tentativa e erro, ao testar diversos protocolos sem embasamento científico, prejudicando a produção. Nesse contexto, a fim de atender essa demanda do setor, este trabalho propõe a Plataforma Crevettic, uma solução focada na gestão de conhecimento e inteligência de negócio das fazendas de camarão, apresentando indicadores aos produtores por meio de análises de dados utilizando técnicas de mineração de dados e aprendizado de máquina. Como prova de conceito foram treinados dois modelos de predição usando floresta randômica. Um voltado para prever o peso médio do camarão no tanque, apresentando coeficiente de determinação de 94%. Outro voltado para prever a ração necessária ideal no tanque, chegando a apresentar 98% no coeficiente de determinação. Essas informações são apresentadas no *dashboard* da plataforma. A plataforma foi desenvolvida usando a arquitetura de microserviços (dividida em módulos) e está sendo utilizada atualmente por sete fazendas parceiras em vários estados do país. A proposta da plataforma Crevettic passou por vários editais de pré-aceleração e aceleração, evoluindo para um produto comercial, que posteriormente se transformou em uma *startup*.

Palavras-chave: Carcinicultura. Sistema de Apoio a Decisão. *Business Intelligence*. Análise de Dados.

ABSTRACT

Farm shrimp production is an important sector in the economy of Northeastern Brazil, mainly in the states of Ceará and Rio Grande do Norte, which together produce 66% of annual national production, about 27.2 thousand tons in total. However, the emergence of new diseases has forced producers to carry out more effective monitoring in relation to the management and health of shrimps. Several information is collected in order to previously identify signs that may cause losses of production, such as slow development and increased mortality rate. However, many farms still control these data inefficiently, using paper or spreadsheets that grows considerably, making them difficult to analyze for humans. Thus, decision-making on farms ends up being based on trial and error, by testing various parameters without scientific basis. In this context, in order to fill out this sector demand, this work proposes the Crevettic platform, a solution focused on knowledge management and business intelligence on shrimp farms, presenting indicators to producers through data analysis using data mining and machine learning. As a proof of concept, two prediction algorithms were trained. First one aimed at predicting the average weight of the shrimp in the tank, with a determination coefficient of 94%. Second one aimed at predicting the ideal necessary ration in the tank, reaching 98% in the coefficient of determination. These information is presented on the platform's dashboard. The platform was developed using the microservice architecture (divided into modules) and is currently being used by seven peaceful farms in various states in the country. The Crevettic platform proposal went through several pre-acceleration and acceleration notices, evolving into a commercial product, which later became a startup.

Keywords: Farming shrimp. Decision Support System. Business Intelligence. Data analysis.

LISTA DE FIGURAS

Figura 1 – Etapas do processo de descoberta de conhecimento em base de dados .	21
Figura 2 – Gráfico caracterizando os portes das fazendas	32
Figura 3 – Principais dados coletados pelas fazendas	33
Figura 4 – Ferramentas usadas pelas fazendas para controle de dados	33
Figura 5 – Satisfação dos produtores de camarão com as soluções adotadas por eles	34
Figura 6 – Cenário da plataforma Crevettic	36
Figura 7 – Arquitetura da plataforma Crevettic	37
Figura 8 – Distribuição de ocorrências por estado	41
Figura 9 – Período das ocorrências	42
Figura 10 – Tipo de ocorrência e faixa etária dos animais	42
Figura 11 – Arquitetura do MVP da Plataforma Crevettic	43
Figura 12 – Tela de relatório de ocorrência do MVP	44
Figura 13 – Esquema da metodologia Scrum	45
Figura 14 – Tela de trabalho do projeto no Trello	46
Figura 15 – Sequência de passos aplicados, seguindo a abordagem do KDD	48
Figura 16 – Exemplo de medição de biometria errada	51
Figura 17 – Heurísticas de manejo mapeadas nas entrevistas	53
Figura 18 – Ilustração Genérica da Validação Cruzada	56
Figura 19 – Histórico de Lançamentos da Plataforma Crevettic	60
Figura 20 – Telas da primeira versão da plataforma Crevettic	60

LISTA DE TABELAS

Tabela 1 – Tabela Comparativa das Soluções x Plataforma Crevettic	31
Tabela 2 – Dicionário de Dados do <i>Dataset</i> Gerado	50
Tabela 3 – Resultados dos testes por algoritmo (Peso Médio Estimado)	58
Tabela 4 – Resultados dos testes por algoritmo (Ração Necessária)	58

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Motivação	15
1.2	Justificativa	16
1.3	Problemática	16
1.4	Objetivo	16
1.4.1	<i>Objetivos específicos</i>	<i>17</i>
1.5	Relevância	17
1.6	Organização do Trabalho	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Sistemas de Apoio à Decisão	19
2.1.1	<i>Data Warehouses e Data Marts</i>	<i>19</i>
2.1.2	<i>Big Data e Cloud</i>	<i>20</i>
2.1.3	<i>Business Intelligence</i>	<i>20</i>
2.2	Descoberta de Conhecimento em Bases de Dados	21
2.3	<i>Data Mining</i>	<i>21</i>
2.4	Inteligência Artificial	22
2.4.1	<i>Aprendizado de Máquina</i>	<i>22</i>
3	CARCINICULTURA	24
3.1	Etapas do Cultivo	24
3.1.1	<i>Maturação e larvicultura</i>	<i>24</i>
3.1.2	<i>Berçário</i>	<i>25</i>
3.1.3	<i>Engorda</i>	<i>25</i>
3.2	Tipos de Cultivo	25
3.2.1	<i>Sistema extensivo</i>	<i>26</i>
3.2.2	<i>Sistema semi-intensivo</i>	<i>26</i>
3.2.3	<i>Sistema intensivo</i>	<i>26</i>
3.3	Considerações sobre o Setor	27
4	O PROGRAMA DE SAÚDE NAS FAZENDAS DE CAMARÃO . .	28
4.1	Biopatometrias	28
4.2	Análises a fresco	28

4.3	Banco de Dados	29
5	TRABALHOS RELACIONADOS	30
6	VALIDAÇÃO DO PROBLEMA E ANÁLISE DE REQUISITOS .	32
6.1	Formulário Online	32
6.2	Entrevistas	34
6.3	Aspectos Econômicos	35
7	PLATAFORMA CREVETTIC	36
7.1	Arquitetura Geral	36
7.1.1	<i>Módulo de ocorrência</i>	38
7.1.2	<i>Módulo de qualidade do ambiente</i>	38
7.1.3	<i>Módulo de biometria</i>	38
7.1.4	<i>Módulo de insumos</i>	39
7.1.5	<i>Módulo de treinamento</i>	39
7.1.6	<i>Módulo de saúde</i>	39
7.1.7	<i>Módulo de business intelligence</i>	39
8	METODOLOGIA DE DESENVOLVIMENTO	41
8.1	Protótipo	41
8.2	Mínimo Produto Viável	43
8.3	Entrevistas e Acompanhamento de Testes Contínuos	44
8.4	Scrum	45
8.4.1	<i>Ferramenta de gerenciamento Trello</i>	46
9	MÓDULO DE BI (BUSINESS INTELLIGENCE)	47
9.1	Levantamento e Integração de Dados	47
9.2	<i>Dataset Gerado</i>	49
9.3	Preparação de Dados	49
9.4	Enriquecimento de Dados	51
9.4.1	<i>Biomassa Estimada</i>	51
9.4.2	<i>Conversão Alimentar</i>	52
9.5	<i>Ranking</i> de Produtividade	52
9.6	Protocolos de Manejo	52
9.7	Correlação dos Dados	53
9.8	Modelos de Predição	54

9.8.1	<i>Peso médio dos animais no tanque</i>	54
9.8.2	<i>Quantidade de ração necessária para cada tanque</i>	54
9.9	Algoritmos Utilizados	54
9.9.1	<i>Treinamento e Testes</i>	55
9.9.2	<i>Métricas de Comparação</i>	56
9.10	Resultados dos Testes	57
9.10.1	<i>Peso Médio Estimado</i>	57
9.10.2	<i>Ração Necessária</i>	58
10	EVOLUÇÃO DA PLATAFORMA	60
11	CONCLUSÕES E TRABALHOS FUTUROS	62
12	PRODUTIVIDADE ACADÊMICA E CONQUISTAS	63
12.1	Trabalhos publicados	63
12.2	Trabalhos apresentados	63
12.3	Capítulo de livros	63
12.4	Participações em palestras e <i>lives</i>	64
12.5	Conquistas	64
	REFERÊNCIAS	65

1 INTRODUÇÃO

A produção de camarão, especialmente da espécie *Litopenaeus Vannamei* ou camarão branco do pacífico (BOONE, 1931), é um importante setor na economia do Nordeste do Brasil, principalmente no Ceará e Rio Grande do Norte, maiores produtores nacionais do crustáceo. A produção nos dois estados corresponde a 29% e 37% da produção nacional, respectivamente. Juntos, os dois estados produzem cerca de 27,2 mil toneladas ao ano, somando aproximadamente 856,9 milhões de reais no comércio (IBGE, 2019). Quando somado aos outros estados do nordeste, esse valor chega a mais de 99%. No cenário Internacional, a *Food and Agriculture Organization of the United Nations* (FAO) coloca o Brasil como um dos maiores produtores de pescados do mundo, destacando-se na América Latina, junto ao México e Equador (FAO, 2012).

1.1 Motivação

Apesar do sucesso, o setor de carcinicultura vem, nos últimos anos, sofrendo grandes impactos devido ao surgimento de novas doenças, como o vírus da mionecrose infecciosa (*Infectious Myonecrosis Virus* (IMNV)), que atingiu principalmente os estados do Nordeste, e o vírus da mancha branca (*White Spot Syndrome Virus* - WSSV), afetando inicialmente o Sul do país e posteriormente espalhando-se também para o Nordeste (COSTA *et al.*, 2011). Essas doenças infecciosas têm causado preocupação, pois podem espalhar-se com facilidade e contaminar um viveiro de camarão inteiro em poucos dias, ameaçando a produção. Em alguns casos, principalmente em áreas com fazendas de micro porte, a produção chegou a cair 90% (REZENDE; MATAVELI, 2017).

Devido ao aparecimento dessas doenças, os produtores têm realizado um acompanhamento mais efetivo em relação ao manejo e saúde dos camarões. Durante esse acompanhamento são aferidos dados sobre a qualidade da água, temperatura e informações biofísicas do animal, entre outras, a fim de identificar previamente sinais que possam causar prejuízos de alguma natureza à produção, incluindo desenvolvimento lento e aumento da taxa de mortalidade, bem como a própria qualidade do crustáceo (NEVES, 2018). Dependendo da gravidade do problema, estes dados são analisados por cuidadores, técnicos ou especialistas para auxiliar no processo de tomada de decisão.

1.2 Justificativa

Contudo, apesar dos cuidados descritos acima, muitas fazendas, principalmente de micro e pequeno porte, ainda realizam o controle desses dados de forma arcaica por meio de planilhas simples em *Excel* ou até mesmo usando papel, dificultando o processamento e análise dos dados. Portanto, este procedimento não é suficiente pois os dados coletados podem aumentar consideravelmente, dificultando o processo para seres humanos. Desse modo, a tomada de decisão nessas fazendas acaba sendo realizada com base na tentativa e erro, ao testar diversas estratégias sem embasamento científico. Além disso, o conhecimento produzido muitas vezes acaba se perdendo com os profissionais.

1.3 Problemática

A partir de entrevistas realizadas durante visitas aos produtores, constatou-se que os sistemas informatizados disponíveis não atendem às necessidades dos produtores, que buscam ferramentas capazes de auxiliar na tomada de decisão. A maioria das soluções disponíveis no mercado focam apenas na gestão financeira e produtiva das fazendas, deixando de lado informações de manejo e saúde do camarão, indispensáveis na tomada de decisão. Além disso, cada produtor desenvolve e aplica sua própria metodologia, que nem sempre se encaixa nos parâmetros usados pelos sistemas existentes.

Contudo, desenvolver uma ferramenta computacional capaz de correlacionar dezenas de parâmetros a fim de auxiliar na tomada de decisão é um desafio. Para tanto, é necessária a criação de mecanismos inteligentes para realizar análises em grandes quantidades de dados, identificando padrões no manejo (incluindo fatores nutricionais e bioquímicos) e identificando possíveis problemas antecipadamente, evitando prejuízos e aumentando a produtividade nas fazendas de camarão.

1.4 Objetivo

A partir da análise preliminar de visitas realizadas a fazendas de camarão, pôde-se obter um entendimento mais amplo das reais necessidades dos produtores a fim de modelar um sistema que atenda a realidade das fazendas de camarão da região (leste cearense e oeste potiguar). Assim, este trabalho tem como principal objetivo propor e desenvolver uma plataforma de *business intelligence* para fazendas de camarão, integrando produtores,

técnicos e especialistas da área para auxiliar na tomada de decisão, minimizando os prejuízos e melhorando a produtividade dessas fazendas.

1.4.1 Objetivos específicos

- I. Analisar o estado da arte sobre sistemas computacionais de auxílio à tomada de decisão no contexto de carcinicultura;
- II. Entender o funcionamento das fazendas de camarão e a importância de seus dados;
- III. Levantar requisitos funcionais e não funcionais a fim de definir o escopo da plataforma proposta;
- IV. Desenvolver um produto mínimo viável como prova de conceito da solução proposta;
- V. Recuperar dados de fazendas de carcinicultura, analisá-los e treinar modelos de predição.

1.5 Relevância

Esse projeto conta com a parceria do Departamento de Ciências Animais (DCA) da Universidade Federal Rural do Semi-Árido (UFERSA), por meio do professor Pedro Martins, além da cooperação das empresas Carcinicultura Kammaros S/A, que produz camarão e tilápia, e PROAQUA JR, que presta serviço de análise da saúde do camarão. Esta pesquisa também é apoiada pelo Projeto de Saúde nas Fazendas de Camarão (PSF Camarão), existente desde 2009, que tem acompanhado as fazendas de camarão da região (Ceará e Rio Grande do Norte), com foco na saúde dos animais.

Mediante estas parcerias, este trabalho conta com uma equipe com profissionais experientes na área da carcinicultura, além do acesso aos dados coletados pelo PSF Camarão. Tais informações, junto às necessidades levantadas por produtores parceiros e por técnicos, dão subsídio ao desenvolvimento de ferramentas de *Business Intelligence* (BI) que agregam valor ao vários níveis do setor, potencializando a tomada de decisão inteligente.

A proposta da plataforma Crevettic foi contemplada em importantes editais de fomento, como o programa Centelha (MCTI). O projeto também teve a oportunidade de participar de vários programas de aceleração, entre eles o Sebrae CE. Além disso, o essa pesquisa ainda produziu diversos artigos acadêmicos.

1.6 Organização do Trabalho

O restante deste trabalho está organizado da seguinte forma. O Capítulo 2 descreve a fundamentação teórica dos conceitos e tecnologias utilizados nesta pesquisa. Já o Capítulo 3 contextualiza área da carcinicultura e seus aspectos, enquanto o Capítulo 4 apresenta a metodologia do Programa de Saúde nas Fazendas de Camarão. O Capítulo 5 destaca os principais trabalhos desenvolvidos no contexto de sistemas de apoio a tomada de decisão na área de carcinicultura. O Capítulo 6 apresenta uma pesquisa de mercado realizada a fim de validar o problema. O Capítulo 7 apresenta a solução proposta, sua arquitetura e funcionamento. No Capítulo 8 é apresentada a metodologia utilizada nessa pesquisa, descrevendo todos os passos realizados até sua conclusão. O Capítulo 9 apresenta e discute os resultados dos testes realizados na pesquisa, destacando os melhores resultados para a solução do problema em estudo. O Capítulo 10 descreve a evolução, desde de a ideação à versão atual. Finalmente, o Capítulo 11 conclui o trabalho através de uma análise do impacto da solução proposta, evidenciando o método utilizado. Este capítulo também sugere propostas para trabalhos futuros. No Capítulo 12 destaca-se a produção desenvolvida durante a pesquisa, incluindo publicações e conquistas.

2 FUNDAMENTAÇÃO TEÓRICA

Com o passar dos anos, surgiram diversas estratégias e ferramentas computacionais com o objetivo de agregar valor aos dados gerados pelas empresas, que estão cada vez mais densos e amplos. A solução desenvolvida nessa pesquisa está fortemente relacionada a essas estratégias. Portanto, para facilitar o entendimento da solução proposta, neste capítulo é apresentada de forma introdutória os conceitos e tecnologias que servirão de base para o desenvolvimento da Plataforma Crevettic.

2.1 Sistemas de Apoio à Decisão

Com a popularização da Computação surgiram diversos Sistemas de Informações e ferramentas de banco de dados capazes de auxiliar no armazenamento e gerenciamento de informações geradas pelas empresas. Com o tempo, os dados se tornaram cada vez mais importantes e valiosos. Então, os sistemas de banco de dados evoluíram para suportar a crescente demanda por relatórios nos diferentes níveis do negócio, tornando-se mais robustos e confiáveis. Em contrapartida, com o crescimento, expansão e globalização das empresas, surgiram novos desafios e necessidades. A demanda dessas empresas fez surgir os Sistemas de Apoio à Decisão (SAD), que auxiliam gestores e profissionais na tomada de decisão. Os SADs usam diversas tecnologias a fim de nortear esse processo (CECI, 2012; BURSTEIN; HOLSAPPLE, 2008)

2.1.1 *Data Warehouses e Data Marts*

O *Data Warehouses* (DW) surgiu para lidar com a demanda por dados nos diversos níveis das empresas, facilitando a geração de relatórios setoriais e gerenciais por meio de uma arquitetura robusta de integração dados. Depois surgiram os *Data Marts* (DM), que são subconjuntos de bancos de dados sobre assuntos (setores) específicos das empresas para facilitar o acesso aos dados por conta da alta demanda, evitando a sobrecarga dos DW. Essas arquiteturas surgiram para dar suporte ao processamento analítico on-line (*Online Analytical Processing* - OLAP), que permite aos usuários criar relatórios e consultas sob demanda e realizar análises de dados por meio de visualizações detalhadas (TURBAN *et al.*, 2009).

2.1.2 *Big Data e Cloud*

Além dos dados armazenados em bases de dados estruturadas, com o surgimento de novos dispositivos eletrônicos e a popularização da internet, a quantidade de dados gerados tem aumentado vertiginosamente. O *Big Data*, como é conhecido, é o conjunto de todo tipo de informação digital provinda de qualquer fonte e formato, como arquivos, banco de dados, áudios, vídeos, etc. Cada empresa hoje gera uma grande quantidade de dados não estruturados. O *Big Data* tem se tornado uma ferramenta poderosa para as empresas. Hoje já existem diversas ferramentas para extração de informações em fontes de dados não estruturados. Estes dados são posteriormente processados por algoritmos de análise de dados. Normalmente todos esses dados ficam armazenados em *Cloud*, uma arquitetura descentralizada de armazenamento e poder computacional sob demanda.

2.1.3 *Business Intelligence*

Posteriormente, surgiram ferramentas e profissionais especializados em analisar e lidar com esses dados a fim de encontrar tendências, padrões e relacionamentos que podem auxiliar no processo de tomada de decisão. BI representa a habilidade de estruturar, acessar e explorar informações com o objetivo de permitir a percepção de conhecimentos antes ocultos, produzindo melhores decisões.

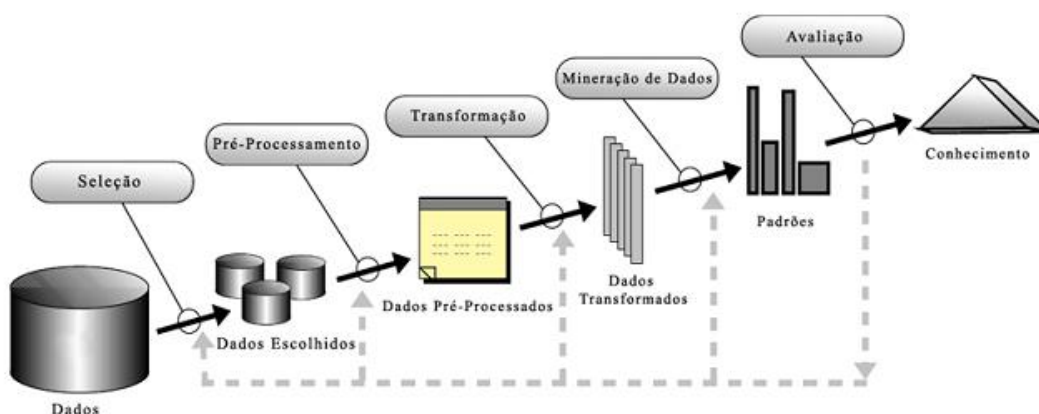
O BI é uma área que une diversos conhecimentos a fim melhorar a tomada de decisão nos negócios. Para tanto, dependendo da necessidade, são analisadas informações de vários setores de uma empresa, a fim de fornecer aos gerentes e analistas de negócios a capacidade de realizar análises mais profundas (TURBAN *et al.*, 2009). Normalmente, todo esse processo é realizado com auxílio de um cientista de dados. O conhecimento sobre o negócio também é indispensável (SANTOS; RAMOS, 2006).

De forma geral, BI são tipos de SAD orientados a dados (BURSTEIN; HOLSAPPLE, 2008). Para tanto, faz uso das mais diversas tecnologias e ferramentas de análise, como descoberta de conhecimento em bases de dados. Os sistemas de BI consistem em coletar informações, registrá-las em bancos de dados e gerenciar o conhecimento que essas informações representam.

2.2 Descoberta de Conhecimento em Bases de Dados

KDD (do inglês *knowledge-discovery in databases*) é a principal metodologia utilizada por cientistas de dados para extrair conhecimento em bases de dados. Para tanto, são aplicadas diversas técnicas e conhecimentos da Computação e Estatística, como Análise e Mineração de Dados (MD) a fim de identificar padrões ou relacionamentos nos dados (FAYYAD *et al.*, 1996). Estas técnicas têm sido amplamente aplicadas por empresas que visam melhorar sua competitividade no mercado. A Figura 1 mostra as principais etapas da metodologia.

Figura 1 – Etapas do processo de descoberta de conhecimento em base de dados



Fonte: Extraída e traduzida de (FAYYAD *et al.*, 1996)

Como representado pela figura, KDD trata-se de uma metodologia iterativa, constituída por diversos passos não triviais para extração de informações implícitas e potencialmente úteis a partir de uma ou várias bases de dados. Para tanto, o conhecimento sobre o contexto do problema é indispensável, sendo o cientista de dados responsável por analisar e interpretar os resultados obtidos ao longo do processo, bem como decidir quais métodos são mais adequados para situação. Algumas etapas devem repetir-se a fim de buscar melhores resultados por meio de sucessivos refinamentos.

2.3 Data Mining

A mineração de dados (*Data Mining* - DM) busca entender e explicar um conjunto de dados a fim de extrair conhecimento. Para tanto, são executados algoritmos a fim de explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação, correlações ou sequências temporais, para detectar relacionamentos

sistemáticos entre variáveis, normalmente invisíveis em análises estatísticas convencionais.

A mineração de dados é um processo extremamente importante para ciência de dados e para metodologia de KDD. Seus procedimentos usam um conjunto amplo de ferramentas e técnicas, incluindo algoritmos de aprendizado de máquina. A mineração de dados combina conhecimentos de estatística e técnicas de inteligência artificial (HAN *et al.*, 2011).

2.4 Inteligência Artificial

A Inteligência Artificial (IA) é uma subárea da Ciência da Computação que estuda formas de reproduzir o raciocínio humano (FACELI *et al.*, 2015). Nos últimos anos foram propostos sistemas computacionais inteligentes capazes de resolver problemas mais genéricos e aprender de maneira autônoma, além de interagirem entre si e com seres humanos. Os sistemas inteligentes podem usar diversas estratégias para solucionar um determinado problema. Algumas abordagens aplicam inferência indutiva a fim de adaptarem-se a novas situações, enquanto outros métodos utilizam modelos matemáticos baseados em probabilidade para buscar conhecimento em grandes conjuntos de dados. Outro método bastante conhecido é o Aprendizado de Máquina (AM), que é uma subárea da IA (AWAD; KHANNA, 2015a).

2.4.1 Aprendizado de Máquina

É uma área de estudo da IA que dedica-se ao desenvolvimento de algoritmos capazes de aprender. Os algoritmos de AM têm sido utilizados no processo de descoberta de conhecimento, pois são capazes de aprender a partir de experiências utilizando reconhecimento de padrões, entre outras técnicas (AWAD; KHANNA, 2015b).

Os algoritmos de aprendizado de máquina usados na mineração de dados podem resolver diversos tipos de problemas, entre eles: classificação, quando busca-se um resultado discreto; regressão, quando busca-se um resultado contínuo; agrupamento, quando busca-se distinguir grupos nos dados. Para tanto, os algoritmos devem passar por um processo de treinamento, através da análise de um conjunto de dados, gerando um modelo de aprendizagem capaz de tratar novas instâncias/situações. Este procedimento pode ser feito de maneira: supervisionada, por meio de exemplos rotulados; não-supervisionada,

por meio dos relacionamentos e similaridades entre os dados; ou a combinação dos dois, conhecida como semi-supervisionada (TAN *et al.*, 2009).

Os algoritmos podem usar diversas estratégias. Alguns algoritmos usam métodos probabilísticos que fazem uso de modelos matemáticos para identificar a disposição dos dados em uma determinada amostra. Os classificadores baseados no teorema de Bayes são exemplos dessa estratégia, utilizada em larga escala em Mineração de Dados. Já os métodos de procura utilizam modelos baseados em árvores para determinar uma descrição hierárquica dos dados. Árvores de decisão e sistemas adaptativos são abordagens que se encaixam nessa estratégia. Outros algoritmos, como os baseados em regras, calculam a similaridade entre as instâncias dos dados (FACELI *et al.*, 2015).

Todas estas abordagens têm evoluído rapidamente e muitos métodos já produzem resultados excelentes em diferentes casos. As metodologias baseadas em AM apresentam ótimos resultados em diversas áreas e são comumente aplicadas no contexto de MD e SAD, bem como no desenvolvimento de BI (STANGE; JOSÉ, 2010).

3 CARCINICULTURA

A carcinicultura é uma subárea da aquicultura voltada para o cultivo de camarões de água doce e salgada. A aquicultura, também conhecida como aquacultura, está relacionada ao cultivo de todo tipo de organismo aquático, como peixes, ostras, mexilhões, algas, etc. O cultivo de camarões é uma atividade antiga, exercida há muitos séculos.

No Brasil, a atividade comercial teve início na década de 70 no Rio Grande do Norte. Embora existam espécies nativas, a maior parte do camarão produzido em cativeiro no Brasil é da espécie *Litopenaeus vannamei* (ou camarão branco do pacífico), originária do Caribe, devido sua forte adaptação às condições de cultivo entre outras características. A produtividade do país hoje está entre as maiores do mundo (SENAR, 2016).

3.1 Etapas do Cultivo

O cultivo do camarão passa por diversas etapas (ou fases), que exigem cuidados diferentes e manejo específico. Portanto, são coletados diferentes parâmetros em cada etapa. Entre elas estão maturação, berçário e engorda, melhor descritas a seguir.

3.1.1 Maturação e larvicultura

A produção de pós-larvas é normalmente realizada em laboratórios (galpões fechados). O processo é, em geral, subdividido em dois estágios independentes: maturação e larvicultura. A maturação é o setor responsável pelo acasalamento e desova. Machos e fêmeas (também conhecida como matrizes) são mantidos juntos, em tanques apropriados, até que ocorra o acasalamento. Após o acasalamento, as fêmeas ovadas são transferidas para os tanques de desova, retornando, posteriormente, aos tanques de maturação.

Os náuplios, animais recém-eclodidos, são estocados em tanques de larvicultura, permanecendo neste setor até atingirem o estágio de pós-larva. A larva do animal passa pelos estágios de náuplio, *protozoa* e *misis* até chegar ao pós-larva e então ir para a próxima fase, que pode ser em berçários ou viveiros de engorda. Até lá, esses animais são mantidos em ambiente altamente controlados.

A fase de pós-larva ainda pode ser dividida em várias etapas, conhecidas como PL 1, PL 2... e assim por diante. O valor da pós-larva também varia conforme sua idade. Cada produtor adquire suas pós-larvas conforme sua necessidade. Normalmente, fazendas que

não contam com berçários optam por pós larvas mais adultas, acima de PL 10 (SENAR, 2016).

3.1.2 Berçário

Os berçários, também conhecidos como berçário intensivo ou pré-berçário, são tanques onde os animais passam as primeiras etapas de vida até sua fase juvenil, quando são levados para viveiros de engorda. Trata-se de uma fase intermediária e opcional. Algumas fazendas dividem esse processo em duas etapas, pré-berçário e berçário, conhecido como trifásico.

Os animais permanecem nesses tanques por um período que varia de 10 a 30 dias, reduzindo assim o tempo nos viveiros. Geralmente, os animais não costumam ficar em berçários após atingirem PL 20. O uso de berçários acarreta em mais custos ao produtor, então nem sempre é adotado, apesar de garantir uma maior sobrevivência (SENAR, 2016).

3.1.3 Engorda

Esta é a última fase da cadeia produtiva do camarão. É nela que os camarões são alimentados até serem despescados e comercializados. É a fase mais longa do processo. O tempo que os animais passam nos viveiros de engorda depende do objetivo comercial dos proprietários ou de fatores inesperados. Nesta fase, os animais são mantidos em espaços maiores e, dependendo do tipo de sistema de cultivo, expostos ao sol e outras influências naturais (SENAR, 2016).

3.2 Tipos de Cultivo

Existem diversas formas de cultivar camarão, conhecidas como sistemas de cultivo. Cada sistema exige um nível de cuidado, aparatos mecânicos e espaços físicos específicos. Antigamente, o cultivo do crustáceo era realizado de forma extensiva, de forma mais natural, sem muitos custos. Com a evolução a partir dos anos, foram surgindo novas formas de cultivar camarões. Entre as mais conhecidas no cenário nacional estão os sistemas extensivo, semi-intensivo e intensivo.

3.2.1 Sistema extensivo

São sistemas de produção mais próximos do ambiente natural de sobrevivência. Inicialmente o cultivo do crustáceo era realizado de forma extensiva, em viveiros abastecidos pelas marés, muitas vezes em conjunto com outras espécies marinhas (cultivo consorciado). Os animais marinhos predadores que por ventura invadem o viveiro são eliminados. Ainda hoje, grande parte das fazendas brasileiras adotam esse método.

Possui baixa densidade (de 0,5 a 4 camarões/ m^2) e baixa renovação de água (2 a 5% do volume). Esse sistema é bastante simples, não utiliza maiores tecnologias, aeradores ou alimentação artificial (ração), sendo que a alimentação baseia-se na produção natural. Normalmente é empregado em fazendas de grandes extensões de terra e viveiros muito grandes (10-20 ha/viveiro). A produtividade gira em torno de 0,2 a 0,7 tonelada/ha/ano (SENAR, 2016).

3.2.2 Sistema semi-intensivo

Utiliza uma densidade de 6 a 20 camarões/ m^2 e renovação de água mais elevada que no sistema anterior (2 a 5% do volume). O uso de aeradores para manter os níveis de oxigênio dentro das exigências da espécie é obrigatório. Devido à maior densidade demográfica, a alimentação baseia-se principalmente no alimento artificial (ração), embora a alimentação natural ainda tenha papel significativamente importante. Normalmente os viveiros têm tamanho entre 1 e 10 ha. Devido às características do solo e estuário, o Brasil adota principalmente o sistema semi-intensivo de cultivo.

3.2.3 Sistema intensivo

Também conhecido como superintensivo, baseia-se em tanques pequenos, densidade de 20 a 120 camarões/ m^2 , aeradores, alta tecnologia, alta taxa de renovação de água (50 a 100% ou mais) e importância extrema do alimento artificial, que deve ser de alta qualidade. Trata-se de um sistema muito produtivo mas também muito arriscado, pois uma leve mudança de parâmetros pode causar grandes prejuízos.

3.3 Considerações sobre o Setor

A produtividade do camarão é influenciada por fatores que vão desde ambiente, sistema de cultivo, clima e tecnologias utilizadas. Isso torna o setor amplo e cheio de procedimentos complexos. Portanto, não é trivial desenvolver uma solução computacional que atenda todo o setor, principalmente porque as decisões tomadas no manejo dos animais podem influenciar outros setores das fazendas, incluindo a comercialização. Entre os principais desafios do setor estão as doenças, principalmente os vírus, que podem dizimar populações inteiras em poucos dias.

4 O PROGRAMA DE SAÚDE NAS FAZENDAS DE CAMARÃO

O Programa de Saúde nas Fazendas de Camarão, ou PSF Camarão como é conhecido, nasceu em 2009 com a missão de produzir e transmitir conhecimento sobre as enfermidades no camarão, principal desafio dos produtores. Idealizado pelo professor Pedro Martins (UFERSA), o programa adota uma metodologia baseada na promoção da saúde (NEVES, 2018).

O programa trabalha em conjunto com os proprietários e técnicos das fazendas, buscando melhores práticas de gestão relacionadas à saúde do camarão cultivado, diminuindo os riscos de aparecimento de doenças. Sua metodologia de trabalho considera fatores relacionados às condições ecológicas e sanitárias durante o cultivo como manejo alimentar, manejo de água, manejo populacional, instalações, equipe técnica e às condições ambientais e tecnológicas da fazenda.

Para tanto eram realizados encontros, *workshops* e visitas técnicas a fim de coletar dados sobre a saúde dos animais nas fazendas. Durante as visitas técnicas eram coletadas informações sobre as ocorrências de problemas nas fazendas. Essas informações eram analisadas a fim de determinar possíveis causas e ações corretivas. Para tanto, foram desenvolvidas ou adaptadas várias ferramentas de análises, entre elas as biopatometrias e análises a fresco.

4.1 Biopatometrias

A biopatometria é um procedimento proposto por (LIGHTNER *et al.*, 1996) que consiste em capturar animais dos viveiros para a realização de um exame que busca a presença de sinais clínicos de enfermidades e seu grau de severidade. O procedimento é realizado durante a biometria, normalmente já feita pelas fazendas. O procedimento tem a função de formar uma ideia mais aproximada da proporção do problema, qual a porcentagem da população que está afetada, qual a severidade das infecções e que tipo de enfermidade está causando o problema.

4.2 Análises a fresco

Animais com sinais clínicos de enfermidades identificados na biopatometria são selecionados para as análises a fresco. Esta técnica baseia-se na observação em microscópio

de tecidos e partes de camarões afetados a fim de estabelecer um diagnóstico presuntivo. As partes a serem observadas ao microscópio são: hepatopâncreas, brânquias, conteúdo intestinal, músculo esquelético e um pedaço da cutícula, seguindo a metodologia proposta por Morales e Cuéllar (2014).

4.3 Banco de Dados

Os dados gerais das propriedades, ocorrências e condições de saúde dos camarões coletados nas biopatometrias e análises a fresco serviam de insumos para a elaboração de relatórios periódicos que auxiliam nas tomadas de decisão. Para gerenciar essas informações foi desenvolvido um banco de dados local usando a ferramenta *Microsoft Access*. Portanto, tanto a coleta, quando a produção desses relatórios eram feitos de forma manual e limitada ao computador com *Access* instalado.

5 TRABALHOS RELACIONADOS

Devido à relevância do setor, diversos trabalhos propõem soluções computacionais para o melhoramento da carcinicultura. Portanto, a fim de entender os problemas enfrentados e como foram solucionados pela comunidade científica, este estudo selecionou e analisou os seguintes trabalhos.

Shareef e Reddy (2018) destacam a importância da análise de parâmetros da água a fim de identificar anomalias. Eles mostram que o *delay* no transporte da água dos tanques até o laboratório pode alterar os dados. Então, propõem uma solução com *hardware* ligado à Internet sem fio para captura desses dados. A fim de desenvolver a solução mais apropriada para a necessidade dos produtores, foram realizadas uma série de entrevistas e análises de estudo de casos. No trabalho também são discutidos os tipos de sensores usados no contexto de aquicultura, possíveis arquiteturas de redes, destacando os desafios levantados por meio de entrevistas, pesquisas e visitas *in loco*. Ao final, os autores apresentam uma solução de sensor a energia solar e sem fio acoplado a um barco motorizado.

No trabalho de Xi *et al.* (2019) é realizado um estudo de caso sobre o setor de criação de camarão, investigando como a realidade aumentada poderia ajudar os produtores de camarão a otimizar operações diárias. Com base na revisão de pesquisas em aquicultura, entrevistas com pesquisadores e consultas nas seis principais fazendas comerciais australianas, destacou-se que o grande desafio da agricultura aquícola é gerenciar e responder a mudanças nas condições da água. Ao final, propuseram uma solução automatizada para coleta e analisa os dados exibidos através de dispositivos com realidade aumentada. A solução pode ser controlada por meio de gestos, deixando as mãos dos cuidadores livres.

Duy *et al.* (2015) propuseram uma arquitetura baseada em redes sem fio que usa diversos sensores e atuadores para o monitoramento, controle e análise de dados de fazendas de camarão. A solução é focada no monitoramento e automação de alguns processos nas fazendas de camarão, como o acionamento de aerogeradores. A arquitetura usa uma topologia *mesh* e rede GSM (*Global System for Mobile*) para dar dinamicidade ao contexto, facilitando o processo de adição ou remoção de nós na rede. Testes de demanda foram realizados e a partir da análise dos resultados os autores afirmam que a solução com os equipamentos sugeridos é viável.

Piplani *et al.* (2015) desenvolveram um software para o gerenciamento da produção

em fazendas de camarão. O software foi construído com base nas necessidades dos produtores, levantadas por meio de entrevistas, testes com protótipos e uma comissão de especialistas. A solução conta com versões *mobile* e *Web*. Os gráficos desenvolvidos ajudam no processo de tomada de decisão, contribuindo para gestão eficiente das fazendas. A solução foi validada e hoje é usada por diversas fazendas. Esta é a solução que mais se aproxima com a proposta deste trabalho, embora não use mineração de dados, nem descreva em detalhes como os gráficos foram construídos.

Os trabalhos supracitados apresentam boas soluções para o problema de coleta de dados e automação de alguns processos no gerenciamento de fazendas de camarão. No entanto, nenhum deles foca na análise de dados especificamente, que é o principal problema dos produtores, entre outras necessidades não atendidas por parte das soluções, demonstradas em capítulos posteriores. A Tabela 1 mostra uma comparação destes trabalhos com a solução desenvolvida nessa pesquisa.

Tabela 1 – Tabela Comparativa das Soluções x Plataforma Crevettic

Solução	Coleta automática	Parâmetros Suportados	Interface	Escalabilidade	Análise de dados
Shareef e Reddy (2018)	sim	temperatura, turbidez, pH, OD	mobile	hardware	apenas propõe
Xi <i>et al.</i> (2019)	não	não especificado	google glass	hardware	apenas propõe
Duy <i>et al.</i> (2015)	sim	temperatura, pH, OD	desktop	hardware	não propõe
Piplani <i>et al.</i> (2015)	não	temperatura, pH, salinidade	desktop e mobile	software	apenas propõe
Plataforma Crevettic	não	personalizável	web responsiva	software	propõe e implementa

OD - Oxigênio Dissolvido

Fonte: elaborado pelo autor (2020).

Por meio da tabela comparativa, é possível perceber que esta pesquisa dá maior atenção ao processamento e análise dos dados coletados, focando na tomada de decisão, que é o principal problema dos produtores, buscando trazer uma maior contribuição para o setor. Diferente das soluções analisadas, a Plataforma Crevettic propõe e realiza uma análise sobre os dados gerados em fazendas de carcinicultura, treinando modelos de predição para auxiliar produtores.

6 VALIDAÇÃO DO PROBLEMA E ANÁLISE DE REQUISITOS

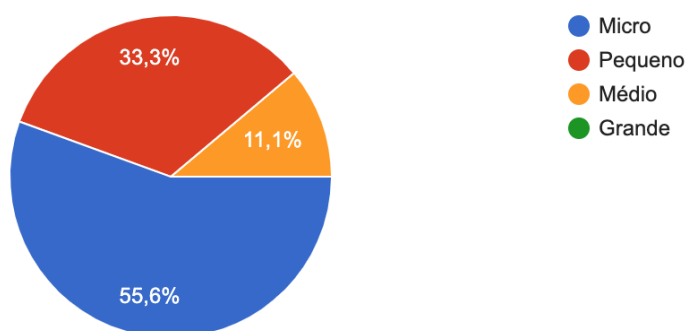
Além de analisar os trabalhos científicos que propõem soluções computacionais para área de carcinicultura, esta pesquisa procurou validar o problema e analisar o perfil das fazendas em relação a adoção de ferramentas para o gerenciamento dos dados gerados. Para tanto, foi realizada uma pesquisa entre os produtores. A pesquisa foi realizada em duas etapas: Formulários Online e Entrevistas com Produtores.

6.1 Formulário Online

Primeiro foi elaborado um formulário para coletar informações sobre como as fazendas lidam com os dados gerados na produção de camarão. Para realizar a pesquisa foi utilizada a ferramenta Google Forms. O formulário foi disponibilizado on-line para uma comunidade técnicos e proprietários de fazendas de Camarão. Ao todo foram obtidas 19 respostas durante um período de um mês.

As primeiras perguntas foram sobre a identificação e classificação da fazenda, como nome, e-mail do responsável, endereço e porte. A maioria das fazendas que participaram do estudo são de micro e pequeno porte, como apresentado no gráfico da Figura 2.

Figura 2 – Gráfico caracterizando os portes das fazendas

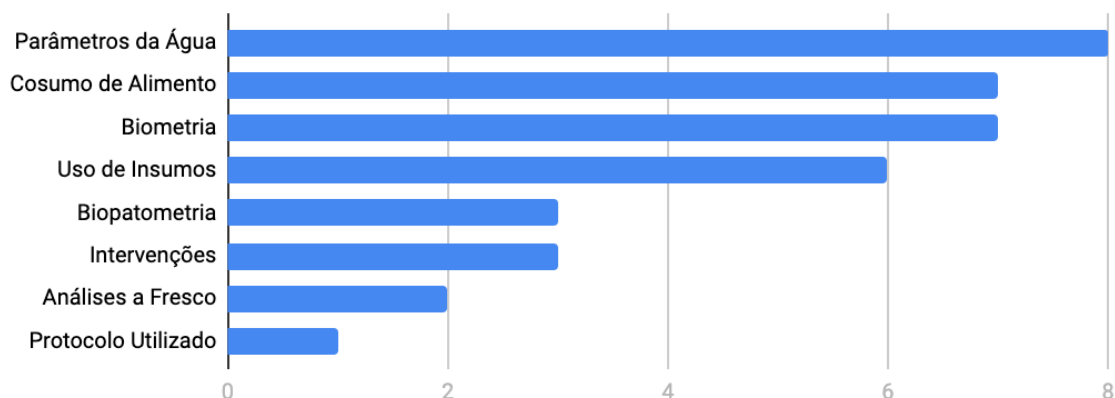


Fonte: Elaborada pelo autor (2020)

Depois o formulário apresenta as seguintes questões: Que tipo de dados a sua fazenda armazena? Com as seguintes opções: parâmetros da água; biometria; biopatometria; análises a fresco; consumo de alimento; consumo de insumos; intervenções. Os formulários ficaram abertos para inclusão de novas opções. A Figura 3 mostra a lista dos dados mais armazenados pelas fazendas.

Depois é questionado como o produtor controla os dados da fazenda, apresentando

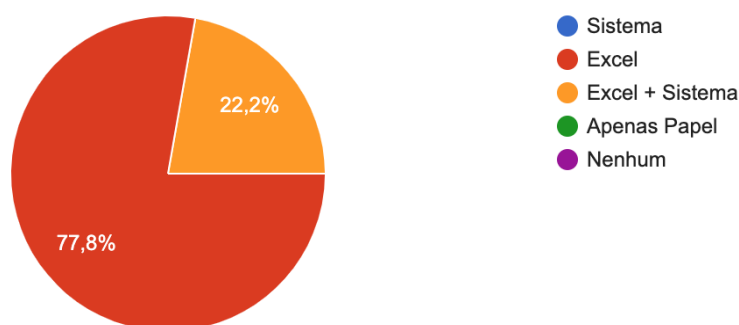
Figura 3 – Principais dados coletados pelas fazendas



Fonte: Elaborada pelo autor (2020)

as seguintes opções: Sistema; Excel; Excel + Sistema; Apenas papel; Nenhum; Outros (aberto). A maioria das fazendas ainda usa apenas Excel, como apresentado pelo gráfico da Figura 4. Contudo, a partir das visitas constatou-se que a maioria das fazendas faz um mal uso da ferramenta, principalmente por falta de profissionais qualificados.

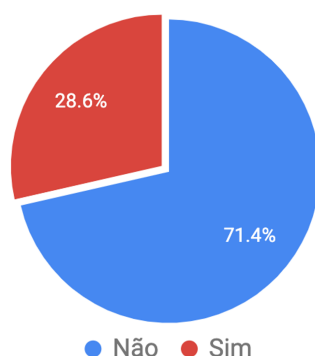
Figura 4 – Ferramentas usadas pelas fazendas para controle de dados



Fonte: Elaborada pelo autor (2020)

Caso a empresa use algum sistema, seja em conjunto com Excel ou não, são coletadas informações sobre as áreas atendidas pelo sistema e suas funcionalidades a fim de entender as necessidades das fazendas. Ao cruzar as informações sobre quais dados são armazenados pelas fazendas com as informações que o sistema adotado contempla, foi possível identificar as principais necessidades do setor, que é, principalmente, o controle de análises de saúde e controle dos protocolos de trabalhos, que são os procedimentos realizados pelos técnicos. Também é questionado se as soluções adotadas atendem as necessidades dos produtores. A maioria dos participantes informou que não, conforme pode ser observado na Figura 5.

Figura 5 – Satisfação dos produtores de camarão com as soluções adotadas por eles



Fonte: Elaborada pelo autor (2020)

Por meio deste estudo também foi possível identificar que existe a necessidade de uma ferramenta para análise dos dados gerados pelas fazendas. Ao perguntar, numa escala de um a quatro, se os produtores acreditam que análises de dados podem revelar ou prever problemas que nem mesmo profissionais experientes conseguem, a maioria respondeu que sim. Além disso, durante uma das entrevistas, um produtor de grande porte relatou que foi necessário desenvolver um *dashboard* próprio para melhorar sua gestão.

6.2 Entrevistas

Foram realizadas entrevistas com quatro produtores. As entrevistas foram realizadas de forma não estruturada, ou seja, não seguiram um roteiro específico. Foram realizadas perguntas abertas, buscando entender a realidade dos produtores locais. Percebeu-se que a maioria das fazendas de micro e pequeno porte terceirizam as análises de saúde dos animais, principalmente por falta de recurso para manter um especialista. Já as necessidades dos grandes grupos de produtores está relacionada à gestão dos insumos e ao custo de produção.

Também foram realizadas entrevistas com especialistas. Segundo eles, a necessidade de controle dos dados de carcinicultura tem aumentado devido à proliferação de novas doenças. Algumas destas doenças, que tem causado grandes prejuízos, apresentam previamente sinais de alerta que podem ajudar a identificar o problema com antecedência.

Além disso, estudos apontam que a eficiência produtiva das fazendas de camarão está relacionada à boa gestão dos dados gerados (ARAÚJO *et al.*, 2018), o que destaca ainda mais a importância de adotar boas ferramentas de gestão de dados.

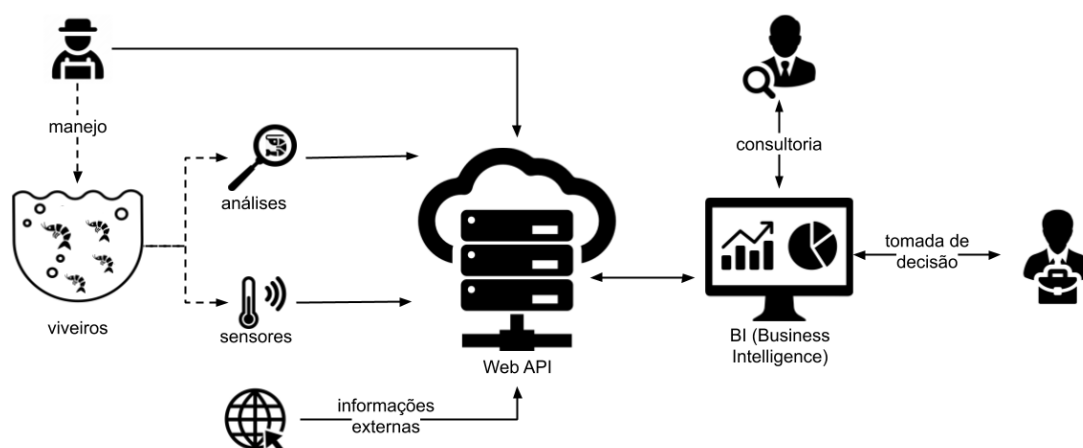
6.3 Aspectos Econômicos

Segundo censo da carcinicultura realizado em 2017, 85% dos produtores do litoral Sul do estado do Ceará pertencem à categoria de micro e pequeno produtor (CEARÁ, 2017). Essa categoria é a que mais sofre com a falta de um sistema que atenda às suas necessidades. A proposta da plataforma Crevettic é focar principalmente nessa categoria, que ainda não possui sistema informatizado e que toma decisões com base apenas em dados de planilhas em Excel ou papel, tornando difícil correlacionar dados de variáveis.

7 PLATAFORMA CREVETTIC

A fim de atender as principais demandas do setor, este trabalho propõe a plataforma Crevettic, uma solução focada na gestão de conhecimento e inteligência de negócio das fazendas de camarão, capaz de oferecer indicadores aos produtores por meio análises de dados que utilizam técnicas de mineração de dados e aprendizado de máquina. Além disso, a plataforma permite a colaboração de especialistas do ramo para tratar problemas específicos. O de uso da plataforma Crevettic é demonstrado na Figura 6.

Figura 6 – Cenário da plataforma Crevettic



Fonte: Elaborada pelo autor (2020)

Como demonstrado na Figura 6, a plataforma recebe os dados: físico-químicos, coletados dos viveiros por cuidadores, técnicos ou através de sensores automáticos; e biofísicos, coletados por especialistas através da análise microscópica dos animais. Os dados podem ser inseridos na plataforma por meio das interfaces Web ou *Application Programming Interface* (API) usando dispositivos móveis, computadores ou por meio da integração da plataforma com outros sistemas. Também é possível importar históricos de coleta, o que agrega valor às análises realizadas. Os componentes de software são melhor descritos na arquitetura da plataforma.

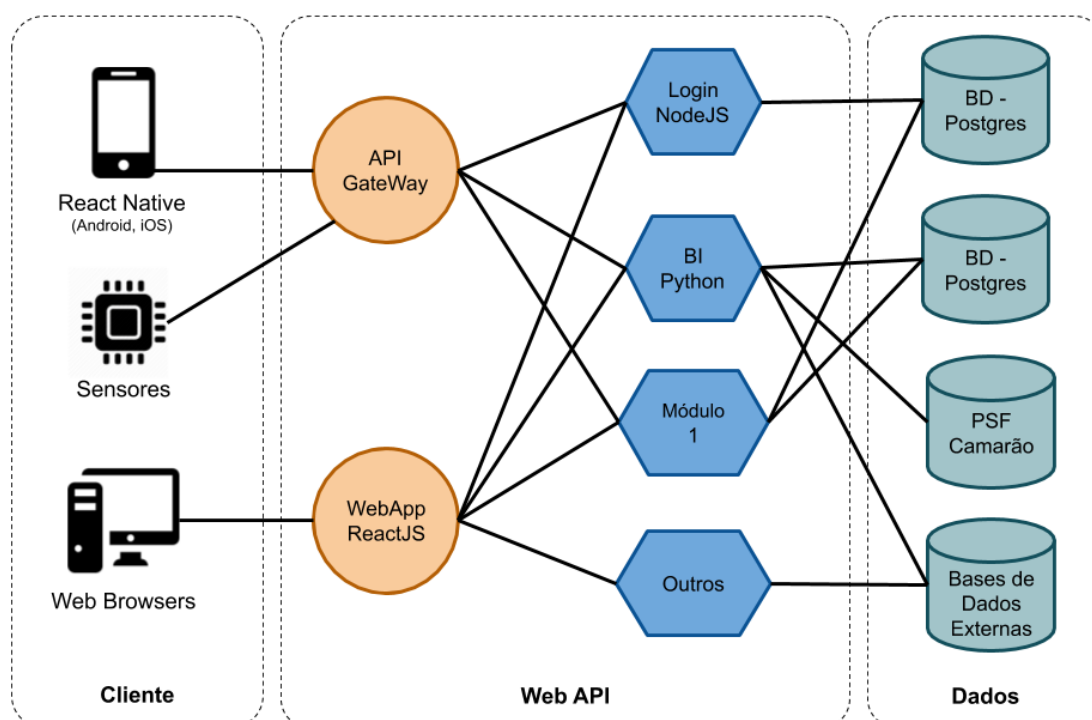
7.1 Arquitetura Geral

A arquitetura da plataforma é baseada em microserviços¹, que facilita a agregação de novas funcionalidades. A arquitetura também permite o uso de componentes/APIs

¹ Modelo de arquitetura desacoplada orientada a serviços

implementados em plataformas e linguagens distintas. Dessa forma, pode haver um maior reaproveitamento de código, além da fácil integração com outros sistemas. Assim, com a arquitetura proposta, é possível desenvolver aplicações voltadas para diversos contextos e níveis de produção, atendendo a diferentes demandas do setor. Como mostrado no capítulo anterior, as necessidades divergem dependendo do porte das fazendas. Na Figura 7 são mostradas as camadas da arquitetura.

Figura 7 – Arquitetura da plataforma Crevettic



Fonte: Elaborada pelo autor (2019)

A arquitetura é dividida em três camadas: (i) Cliente, onde aplicações desenvolvidas em diferentes plataformas podem consumir seus serviços usando o padrão *restfull*; (ii) Web API, onde é realizado todo o controle da plataforma e disponibilizada suas funcionalidades como microserviços; e (iii) Dados, onde são armazenadas e recuperadas informações de diversas bases distribuídas.

A Web API conta com módulos independentes e desacoplados que acessam os dados através da orquestração do módulo de segurança, onde é realizado o controle de usuários, login e permissões de acesso. Os demais módulos são melhor descritos a seguir.

7.1.1 Módulo de ocorrência

Trata-se de um subsistema desenvolvido para coleta de ocorrências (problemas na produção do crustáceo) a fim de gerar uma base de dados de nível local, regional e nacional de modo a apoiar gestores em caso de epidemias ou problemas coletivos. Para tanto, através da plataforma, é possível cadastrar uma nova ocorrência fornecendo informações sobre: o ambiente de cultivo; idade do camarão; parâmetros da água; tipo de problema; estimativa de prejuízo; entre outras.

Estas informações são avaliadas pela equipe do Programa de Saúde nas Fazendas de Camarão. Com estas informações, a equipe pode gerar e disponibilizar relatórios periódicos, facilitando o panorama da carcinicultura. Dependendo da gravidade, são realizadas visitas a fazenda a fim de analisar melhor o problema. Os responsáveis pelo programa têm acesso aos dados de todas as ocorrências. O módulo é disponibilizado gratuitamente para os produtores, gerando uma rede colaborativa de informações.

7.1.2 Módulo de qualidade do ambiente

Esse módulo tem o objetivo controlar as informações do ambiente onde os animais vivem (viveiros). Segundo as informações coletadas nas entrevistas e trabalhos científicos (SOSA *et al.*, 2016), esses parâmetros ambientais influenciam fortemente na qualidade da produção, podendo inclusive, facilitar a proliferação de doenças ou baixo desenvolvimento dos animais. Entre as informações coletadas nas fazendas estão principalmente aspectos físico-químicos da água, acidez (pH), temperatura (°C), transparência e turbidez, entre outros. Como destacado no Capítulo 6, essas são as informações mais usadas pelos carcinicultores. É possível, também, personalizar os parâmetros à realidade de cada fazenda.

7.1.3 Módulo de biometria

Além dos parâmetros ambientais, as fazendas também fazem o acompanhamento dos parâmetros dos animais, principalmente seu peso e tamanho médio. Com essas informações é possível calcular a massa adquirida durante um período de tempo, analisar o desenvolvimento do animal, entre outras coisas. Essas informações são muito importantes para avaliar e prever a produtividade da fazenda como um todo, bem como de viveiros

específicos.

7.1.4 *Módulo de insumos*

É o módulo voltado para controlar entradas de insumos, incluindo ração, insumos químicos, aditivos e biorremediadores, bem como misturas de vários compostos. A carcinicultura é um ambiente de testes, então é necessário ter o controle desses dados a fim realizar análises comparativas de produção. Essas análises podem guiar os gestores na tomada de decisão.

7.1.5 *Módulo de treinamento*

A partir da parceria com o PSF Camarão foi possível construir um módulo educacional voltado para o treinamento de profissionais que trabalham na produção de camarões. São vídeos e cursos com as melhores práticas de manejo e como a plataforma Crevettic pode ajudar a melhorar a produção, gerenciando os dados gerados na fazenda. Os vídeos contam com conteúdos teóricos e práticos produzidos em parceria com especialistas da área.

7.1.6 *Módulo de saúde*

Este módulo busca gerenciar as análises de saúde dos animais na fazenda. Essas análises são realizadas por profissionais qualificados em laboratórios bem equipados. Para tanto os camarões são dissecados a fim de avaliar seus tecidos. Os profissionais, com acesso prévio ao sistema, podem inserir os dados durante o processo. Assim, é possível ter acesso ao resultado das análises remotamente em tempo real, além de acompanhar a evolução de saúde dos animais por meio dos registros anteriores. O módulo de saúde também é fruto de uma parceria com os profissionais do PSF Camarão. A partir dessas análises é possível prever com antecedência possíveis problemas na produção, como doenças, entre outros.

7.1.7 *Módulo de business intelligence*

É o módulo responsável por integrar todas as informações da plataforma, gerando gráficos e relatórios que descrevem padrões de comportamento nos dados gerados pelas fazendas. É nesse módulo que são apresentados os resultados dos cruzamentos de diversos

dados da fazenda, mostrando os relacionamentos entre eles. Assim, é possível auxiliar os gestores no processo de tomada de decisão. Para tanto é realizada a análise dos dados usando técnicas de mineração de dados, a fim identificar padrões que são apresentados de forma visual aos usuários da solução. A contribuição mais relevante deste projeto está nesse módulo, pois agrega grande valor aos dados coletados. No Capítulo 9 é explicado seu funcionamento com mais detalhes.

8 METODOLOGIA DE DESENVOLVIMENTO

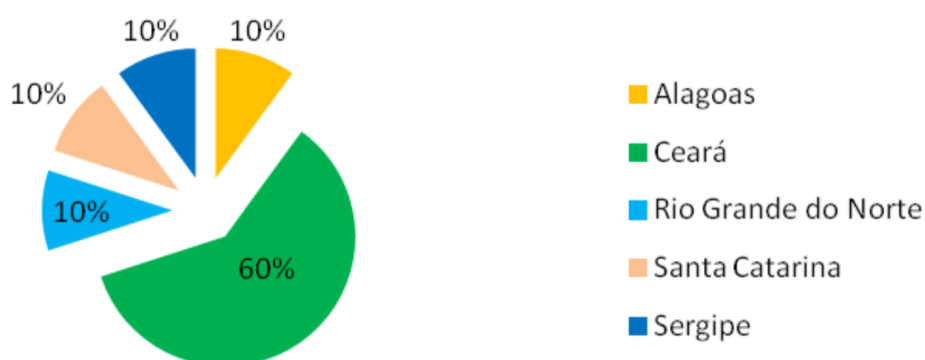
A metodologia seguida pelo projeto é inspirada em algumas metodologias de validação de ideias, prototipação e expansão de produtos. O propósito dessa pesquisa é entregar uma ferramenta (produto) útil que seja amplamente adotada pelas fazendas de camarão. Para tanto, procurou-se validar cada artefato desenvolvido.

8.1 Protótipo

A fim de validar a necessidade do projeto como um todo, foi elaborado um protótipo, usando o formulário do Google, para coletar ocorrências nas fazendas de carcinicultura da região. O formulário tinha 19 campos, entre eles: e-mail; nome; endereço; responsável pela fazenda; ano e mês da ocorrência; grau do problema; quantidade de viveiros em funcionamento; quantidade de viveiros afetados pelo problema; temperatura e salinidade mínima e máxima aferida nos viveiros afetados; peso estimado do camarão; densidade e semana de cultivo dos viveiros afetados; descrição do problema; custo estimado de perda.

O protótipo foi liberado entre os produtores. Ao todo foram coletados 28 ocorrências durante um período de dois meses. A partir desses dados foi gerado um relatório, divulgado publicamente (preservando a identidade das autores das notificações). Os produtores que responderam ao formulário são de fazendas localizadas em cinco Estados. A região Nordeste apresentou a maior participação (90%), seguida da região Sul (10%). O Estado do Ceará apresentou o maior registro de ocorrências (60%), como mostra a Figura 8.

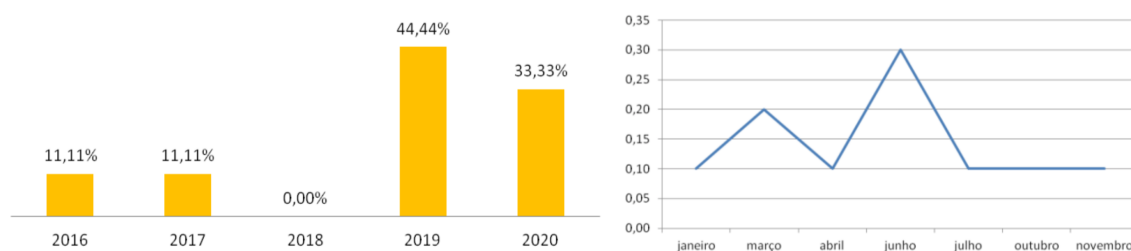
Figura 8 – Distribuição de ocorrências por estado



Fonte: Elaborada pelo autor (2020)

As ocorrências registradas foram em sua maioria referentes aos anos de 2019 (44,44%) e 2020 (33,33%). Nesse boletim foram identificados dois picos mensais, sendo o primeiro no mês de junho (30%) e o segundo no mês de março (20%), conforme ilustrado na Figura 9.

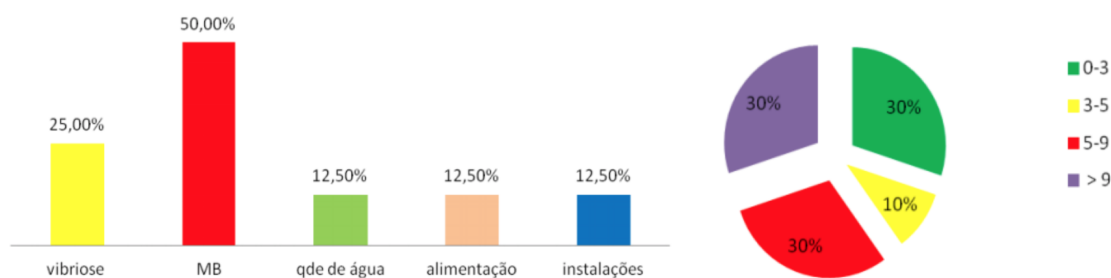
Figura 9 – Período das ocorrências



Fonte: Elaborada pelo autor (2019)

Ao longo dos ciclos o período relatado com maiores problemas foi na 6ª semana (33%) e 8ª semana (33%) de cultivo. Entre os problemas mais frequentes relatados nas ocorrências estão: baixo crescimento; lentidão letárgica; camarão não se alimenta; surto de vibriose; mortalidade e mancha branca. Essas ocorrências na maioria dos casos (60%) atingiram as populações de camarão com peso médio maior do que 5g (Figura 10).

Figura 10 – Tipo de ocorrência e faixa etária dos animais



Fonte: Elaborada pelo autor (2020)

As principais ocorrências tiveram relação com a virose, denominada de mancha branca (MB) (50%), bem com a infecção, geralmente secundária, relacionada a vibriose (25%). Outras ocorrências nesse período tiveram relação com qualidade de água (nitrito), falha no manejo alimentar em virtude de perda de apetite e falhas em instalações, cada uma com uma frequência de 12,5%, , conforme apresentado na Figura 10.

Os casos relatados de ocorrência de mancha branca normalmente estão relacionados à temperatura. Não foram feitas análises de associação entre essas duas variáveis. Entretanto, os valores médios de temperatura foram baixo e, em alguns casos, com grande

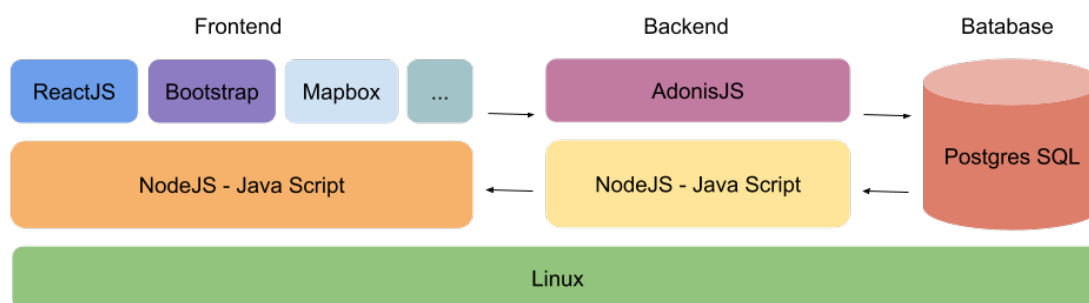
amplitude térmica. Com relação à salinidade foi possível verificar que as fazendas localizadas em baixa salinidade também apresentaram muitas ocorrências.

O desenvolvimento do protótipo seguiu uma metodologia de validação proposta por Knapp *et al.* (2017), engenheiro do Google. Portanto, procurou-se validar a ideia do módulo de ocorrência de forma prática, coletando os dados por meio de um formulário e entregando relatórios estáticos antes de desenvolver a ferramenta propriamente dita.

8.2 Mínimo Produto Viável

A partir dos bons resultados obtidos com o protótipo, pode-se elaborar um Mínimo Produto Viável (*Minimum Viable Product* - MVP) a fim de coletar periodicamente as ocorrências de problemas na produção do camarão, incluindo prejuízos relacionados à mancha branca, principal problema dos produtores. Um MVP é uma versão reduzida e funcional de uma solução com o objetivo de testar sua viabilidade/utilidade no mercado. A Figura 11 mostra com mais detalhes os principais componentes e tecnologias utilizadas no MVP.

Figura 11 – Arquitetura do MVP da Plataforma Crevettic



Fonte: Elaborado pelo autor (2020).

O MVP foi construído usando a linguagem JavaScript para rodar no ambiente Node.js¹. Foram utilizados diferentes *frameworks* no desenvolvimento. Para desenvolver os componentes gráficos e telas da aplicação foi utilizado o ReactJS² junto as várias bibliotecas de componentização e estilização. Já para desenvolver as regras de negócios e controle de acesso foi utilizado o AdonisJS³, um *framework* voltado para o desenvolvimento ágil de aplicações robustas. O banco de dados utilizado foi o PostgreSQL⁴. O

¹ <https://nodejs.org>

² <https://reactjs.org>

³ <https://adonisjs.com>

⁴ <https://www.postgresql.org>

sistema é responsivo, então pode ser acessado tanto em dispositivos móveis quanto em *desktop* através de qualquer navegador Web. A Figura 12 mostra a tela de relatório de ocorrências do MVP.

Figura 12 – Tela de relatório de ocorrência do MVP



Fonte: Elaborado pelo autor (2020).

O MVP desenvolvido permitiu aos produtores cadastrarem suas fazendas, relatar suas ocorrências e visualizar todas as ocorrências cadastradas pelos produtores em forma de relatório. Esse MVP se transformou no Módulo de Ocorrências, que hoje conta com gráficos e mapas espacotemporais dinâmicos, além de possibilitar o gerenciamento de suas ocorrências individualmente.

O MVP foi liberado entre os produtores e passou a substituir o protótipo. A partir daí pôde-se construir um planejamento de curto, médio e longo prazo do projeto, incluindo o lançamento dos outros módulos idealizados para o sistema. O planejamento deixou margem para mudanças, seguindo a máxima da metodologia MVP, que é errar rápido para consertar rápido (RIES, 2012).

8.3 Entrevistas e Acompanhamento de Testes Contínuos

A fim de avaliar a adoção das funcionalidades desenvolvidas, bem como permitir *feedbacks* para melhorias constantes, foram criados grupos de discussões com fazendas parceiras. Essa metodologia permitiu entender melhor as demandas e principais problemas do setor na perspectiva dos fazendeiros. O grupo iniciou com duas fazendas da região leste

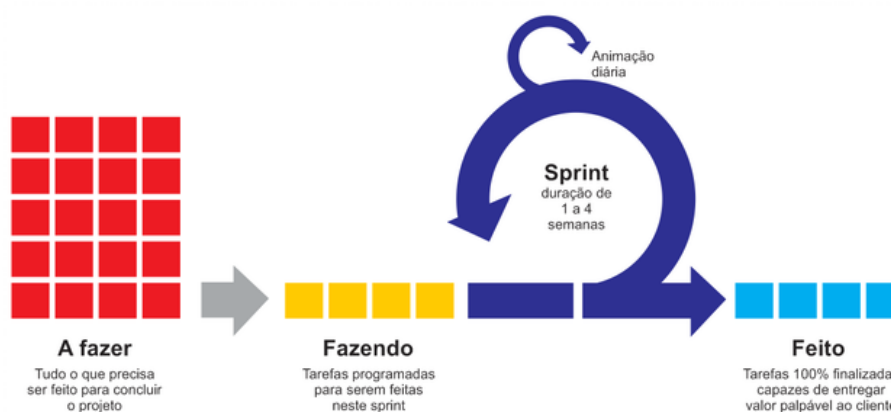
do ceará, posteriormente aumentando para sete (incluindo fazendas de outros estados). Além das primeiras visitas descritas no Capítulo 6, foram realizadas novas entrevistas com profissionais da área, líderes de associações e produtores de camarão de vários portes. Também foram realizadas diversas palestras e *lives* para divulgação da plataforma. Essas atividades fizeram parte da metodologia de desenvolvimento adotada.

8.4 Scrum

Para atender a estratégia de MVP, este trabalho buscou aplicar uma metodologia modular e adaptável focada na entrega rápida de incrementos (artefatos) da solução em cada ciclo de desenvolvimento. Desse modo, os produtores podem usufruir das funcionalidades da solução previamente, antes de concluir o desenvolvimento do sistema inteiro. A arquitetura baseada em microserviços demonstrada no Capítulo 7 também foi pensada para essa estratégia de desenvolvimento.

Para tanto, a metodologia Scrum foi utilizada, pois é voltada para o desenvolvimento ágil de software, que usa um Modelo de Processo Iterativo e Incremental, conforme ilustrado na Figura 13. A metodologia tem sido amplamente utilizada no desenvolvimento de software porque permite um alto grau de modificações no escopo do projeto durante o desenvolvimento (SUTHERLAND, 2016).

Figura 13 – Esquema da metodologia Scrum



Fonte: Extraída do site ergotriade.com.br (2020)

Como mostrado na Figura 13, a metodologia usa ciclos iterativos de desenvolvimento conhecidos como Sprints. Nesse projeto, cada ciclo tem duração de duas semanas. Ao final de cada Sprint, um incremento é implantado no ambiente da fazenda de camarão para ser validado e gerar *feedbacks* para melhoria nas iterações posteriores. A cada Sprint

é realizada uma reunião de equipe, incluindo usuários do sistema, para avaliar os pontos de melhoria e planejar a nova Sprint. Também são realizadas reuniões de alinhamento diárias (*stand-up*) entre os desenvolvedores. Um representante dos criadores de camarão acompanha todo o processo, assumindo o papel de *Product Owner* (PO).

8.4.1 Ferramenta de gerenciamento Trello

Para organizar as atividades da equipe de desenvolvimento, foi inicialmente adotada a ferramenta Jira (da *Atlassian*)⁵. No entanto, devido a limites de uso da conta gratuita, migrou-se para a ferramenta Trello, um software voltado para o gerenciamento de atividades (ver Figura 14).

Figura 14 – Tela de trabalho do projeto no Trello



Fonte: Captura de tela feita pelo autor (2020)

Para adaptar-se a proposta do *Scrum*, foram criadas as listas de: "*Product Backlog*", que representam todas as histórias (necessidades dos clientes); "*Sprint Backlog*", com o que deve ser feito durante o *sprint* atual; "Fazendo", o que já começou a ser desenvolvido; "Em teste", com o que está pronto esperando validação; "Feito", com o que foi concluído e validado para *deploy*. Também foram adicionadas as listas "Com problemas" e "Descartados", que representam as atividades impedidas por algum motivo e as descartadas, respectivamente.

⁵ <https://www.atlassian.com/br/software/jira>

9 MÓDULO DE BI (*BUSINESS INTELLIGENCE*)

Além de desenvolver os módulos clássicos de cadastro e apresentação de dados necessários para o sistema, a principal contribuição deste trabalho são as ferramentas de BI que podem descrever os relacionamentos ou padrões de comportamento nos dados coletados, sugerindo intervenções através de um *dashboard*. Para tanto, é necessário entender previamente a relação desses dados.

Então, paralelamente ao desenvolvimento, foi realizada uma análise de dados que seguiu uma metodologia simplificada inspirada no processo clássico de *Knowledge Discovery in Database* (KDD) (FAYYAD *et al.*, 1996) que usa técnicas iterativas com sequências de passos recorrentes a fim de alcançar resultados mais precisos, como a fase de testes e ajustamento (BRAGA *et al.*, 2018). Para tanto, a estratégia busca selecionar e ajustar os melhores algoritmos de acordo com o contexto estudado. A Figura 15 mostra a sequência de etapas para construção dos modelos de predição para o módulo de BI.

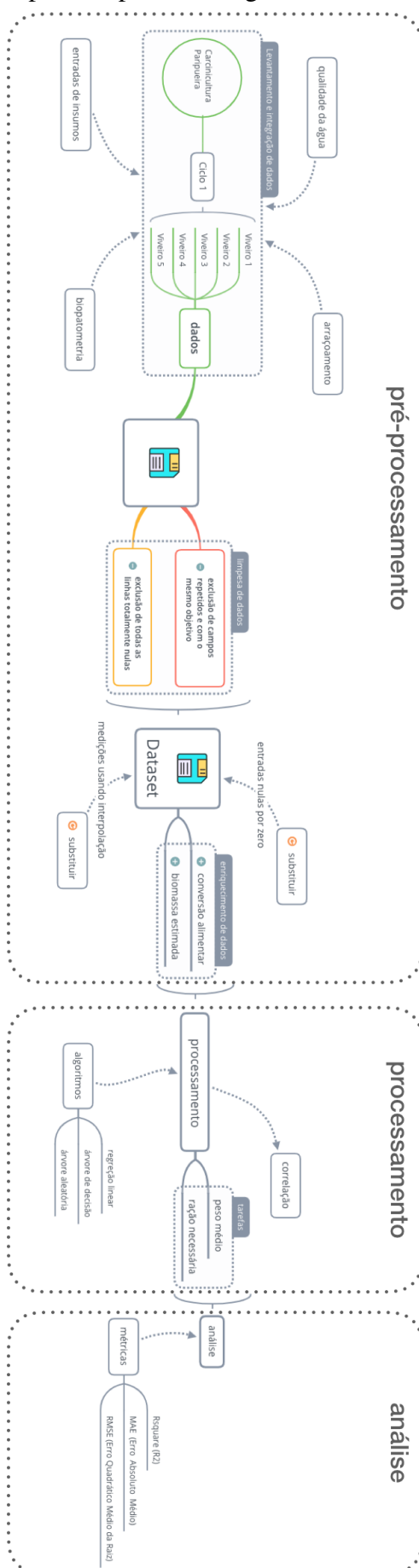
A metodologia pode ser dividida em três fases: **(1) pré-processamento**, onde os dados são levantados, tratados e preparados para serem, então, processados; **(2) processamento**, onde os dados são processados pelos algoritmos, gerando modelos que posteriormente são testados e avaliados; **(3) análise**, onde os resultados dos testes são interpretados e comparados a fim de entender o comportamento dos dados e extrair conhecimento.

9.1 Levantamento e Integração de Dados

O maior desafio desse projeto foi a busca por dados úteis de qualidade. Enquanto eram realizadas as visitas aos produtores, a fim de levantar requisitos e entender o dia-a-dia dos carcinicultores, buscava-se também recuperar dados gerados pelas fazendas de camarão. Contudo, a maioria desses dados havia sido perdida ou estava em pastas físicas difíceis de recuperar.

Os dados recuperados para análise foram obtidos da Fazenda Carcinicultura Paripoeira, situada no Rio Pirangí, na região de Beberibe. Os dados foram coletados em cinco tabelas: Análise da água, incluindo dados sobre parâmetros de qualidade da água e ambiente; Biopatometria, informações sobre crescimento do animal e análises preventivas superficiais; Ração, entradas de alimento nos tanques/viveiros; Biorremediadores, uso de biorremediadores na água e ração; Insumos Químicos, aplicação de produtos na água a fim

Figura 15 – Sequência de passos aplicados, seguindo a abordagem do KDD



Fonte: Elaborada pelo autor (2020)

de balancear parâmetros.

As tabelas estavam estruturadas e os dados organizados em padrões distintos. Os únicos campos comuns eram de data ou dia de cultivo, o que foi usado para correlacionar e integrar as tabelas. A integração ocorreu manualmente por conta da discrepância e falta de padrão nas tabelas. Por exemplo, às vezes a data era representada em uma linha, outras vezes em uma coluna.

9.2 Dataset Gerado

Ao final da integração obteve-se uma tabela com 44 colunas. Para entender melhor cada uma, foi elaborado um dicionário de dados (ver Tabela 2).

Foram recuperados dados de um ciclo de produção, envolvendo 6 viveiros de engorda. Os três primeiros viveiros, com área de 0,45 hectares, foram populados com densidade de 25 PL (pós larvas) por metro quadrado. Já os três últimos, com área de 0,15 hectares, foram populados com densidade de 40 PL (pós larvas) por metro quadrado. A espécie de camarão produzida é *Litopenaeus Vannamei*. Todos os viveiros foram populados dia 13 de novembro de 2019 com larvas na idade de PL 16 diretamente da larvicultura num processo monofásico.

Todas as tabelas, de todos os viveiros, foram integradas em um só *dataset*, contendo 436 observações no total. Os dados reportam até 87 dias de cultivos, mas não há informações sobre despesca. No entanto, havia muitos dados nulos. Para resolver esse problema, evitando excluir linhas e colunas, foram adotadas várias estratégias, apresentadas a seguir.

9.3 Preparação de Dados

Considerou-se os dados de dois tipos: entrada, dados de arrazoamento, biorremediadores, insumos químicos; coleta, dados sobre qualidade da água e biopatometrias. Os dados ausentes de entrada foram preenchidos com zeros, pois correspondem a nenhum valor de entrada. Já os dados de coleta foram preenchidos com uma interpolação linear usando os últimos valores anterior e posterior válidos, por meio da Equação 9.1, pois representam dados faltantes nesse intervalo.

$$\frac{y - y_0}{y_1 - y_0} = \frac{x - x_0}{x_1 - x_0} \quad (9.1)$$

Tabela 2 – Dicionário de Dados do *Dataset* Gerado

Origem	Parâmetro	Descrição	Unidade
Qualidade da água	viveiro	viveiro	-
	dia	dia de cultivo	-
	data	data de referência	-
	salinidade	nível de salinidade da água	%
	ph	potencial hidrogenionico	ph
	sat_od	saturação de oxigênio	%
	oxigenio	oxigenio dissolvido	ppm
	alcalinidade	nível de alcalinidade da água	mg/l
	dt	dureza total	mg/l
	no	nível nitrito	mg/l
	nh	nível amônia	mg/l
	temp	temperatura da água	°C
	secchi	transparência (disco secchi)	centímetro
	cor_agua	tonalidade da água	nominal
	salinidade_tarde	nível de salinidade da água	%
	ph_tarde	potencial hidrogenionico	ph
	sat_od_tarde	saturação de oxigênio	%
	oxigenio_tarde	oxigenio dissolvido	mg/l
	temp_tarde	temperatura da água	°C
	taxa	renovação de água	%
Ração	rac_manha	ração	kg
	rac_tarde	ração	kg
	rac_noite	ração	kg
	racao	ração total diária	kg
Biorremediadores	alliplus	óleo essencial	ml
	pbsaude	probiótico saúde (intestinal)	g
	lts_agua	água da mistura	litro
	puim_arroz	prebiótico puim de arroz	kg
	pbequilíbrio	probiótico equilíbrio (ambiente)	g
	simbioticos	prebiótico + probiótico	litro
Insumos Químicos	calcio	hidróxido	kg
	dolomita	cal dolomita	kg
	nPK	nitrito + potássio + cálcio	kg
	nitrito	fertilizante químico	kg
	potássio	fertilizante químico	kg
	magnésio	fertilizante químico	kg
	lithonutri	algas marinhas	kg
Biopatometria	qtd_camarao	quantidade da amostra	qtd
	crescimento	diferença da última biometria	grama
	ant_verm	antenas vermelhas	%
	hp_branco	hepatopâncreas brancas	%
	peso_total	peso total da amostra	grama
	peso_medio	peso estimado de cada camarão	grama

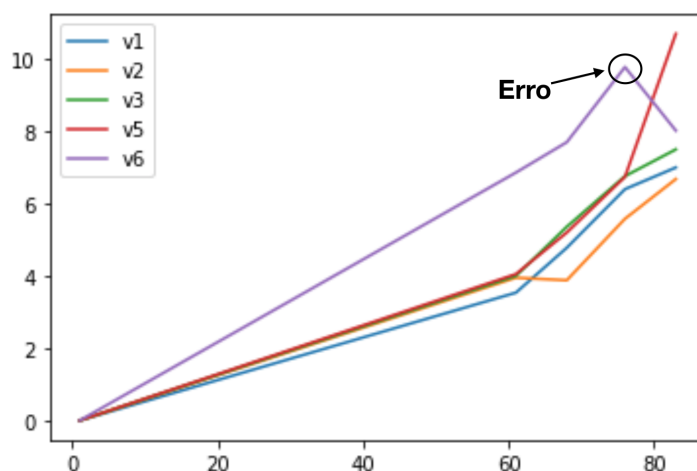
Fonte: elaborado pelo autor (2020).

Com essa estratégia pôde-se manter todas as linhas da tabela, evitando exclusões.

A partir da comparação de crescimento do peso do camarão, percebeu-se que, em algumas biometrias, o camarão parecia diminuir de tamanho, o que não faz sentido, como pode ser visualizado na Figura 16. Depois de algumas entrevistas constatou-se

que o problema poderia ter sido causado por um erro de medição, pois algumas vezes o camarão é posto com muita água na balança. Por conta disso, para evitar problemas no treinamento dos algoritmos, os valores de biometria que apresentavam essa característica foram excluídos. Para evitar que esse problema se repita durante a inserção de dados no sistema, é emitido um alerta sempre que o usuário tenta entrar com uma biometria menor que a anterior.

Figura 16 – Exemplo de medição de biometria errada



Fonte: Elaborada pelo autor (2020)

9.4 Enriquecimento de Dados

Além dos dados de entradas e medições, o *dataset* foi enriquecido com alguns dados obtidos por meio de cálculos entre os atributos existentes. Esses dados representam informações importantes que subsídio ao processo de tomada de decisão, conforme observado durante as visitas.

9.4.1 Biomassa Estimada

A biomassa do viveiro leva em consideração o volume de água no tanque (*Volume*), densidade de animais por m^3 (*Densidade*), peso médio do animal (*Peso*) e taxa de sobrevivência (*Sobrevivencia*). A Equação 9.2 exemplifica esse cálculo:

$$Biomassa = \left(\frac{Peso * Area * Densidade}{Sobrevivencia} \right) \quad (9.2)$$

Com isso é possível estimar a massa total de animais do viveiro. A taxa de sobrevivência é um valor de 0 a 1 que representa o percentual de animais vivos no viveiro. Trata-se de um valor especulativo, adquirido pelo cuidador do viveiro.

9.4.2 Conversão Alimentar

Para avaliar a qualidade da produção do viveiro, é realizado o cálculo da conversão alimentar, que leva em consideração os insumos consumidos (ração, biorremediadores e etc.) pelos animais sobre seu peso médio (*Peso*). A Equação 9.3 mostra um exemplo do cálculo da ração (*Racao*).

$$Conversao = \left(\frac{Racao}{Peso} \right) \quad (9.3)$$

Esse atributo, obtido através do enriquecimento dos dados, é a principal informação para comparar a produtividade do viveiro. Quanto melhor (maior) a taxa de conversão alimentar, mais produtivo é o viveiro, pois um dos maiores gastos de uma fazenda é a ração consumida pelos animais.

9.5 Ranking de Produtividade

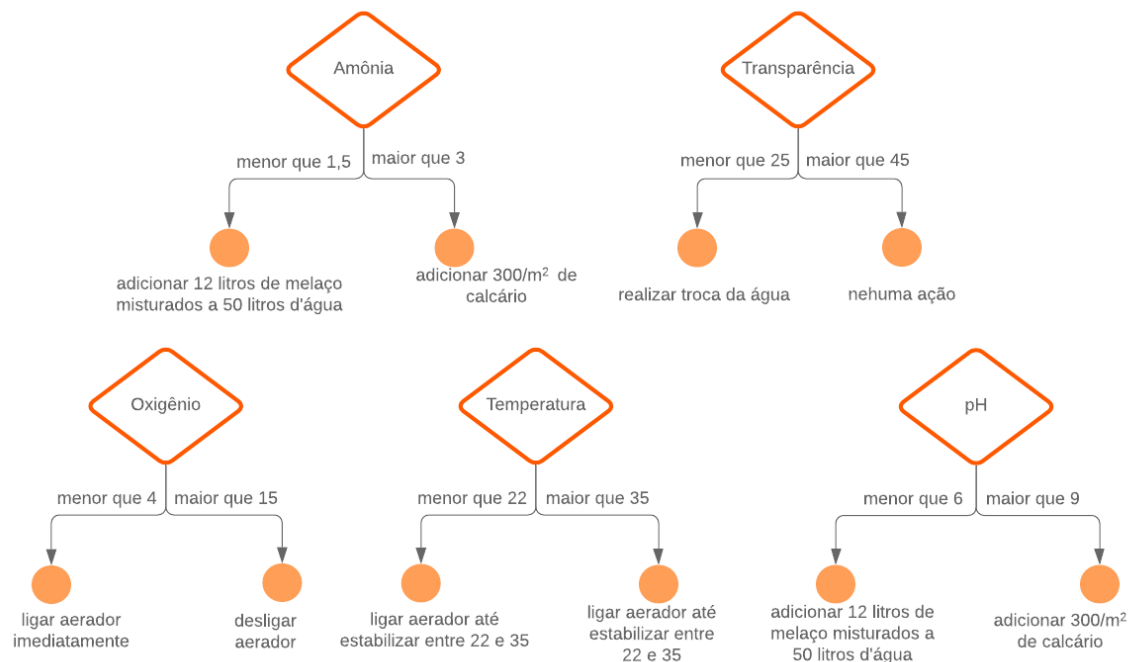
A fim de nortear os gestores e melhorar a produtividade das fazendas, foi desenvolvido o Ranking de Produtividade, que é um comparativo dos melhores viveiros de todas as fazendas cadastradas na plataforma. Para manter segurança, as informações identitárias são ocultas, evitando a identificação das fazendas. Para desenvolver o *ranking* foi usado como base o cálculo de conversão alimentar.

9.6 Protocolos de Manejo

Segundo levantamento realizado em algumas fazendas, o manejo dos viveiros segue algumas regras pré-estabelecidas (protocolos) que se baseiam em estudos científicos e observações práticas. A partir do acompanhamento das fazendas pôde-se mapear alguns desses protocolos de manejo. A Figura 17 mostra suas heurísticas.

A partir dessas regras foi desenvolvido um sistema de notificações para auxiliar técnicos a realizar o manejo corretamente, evitando problemas mais graves. Outra ferramenta

Figura 17 – Heurísticas de manejo mapeadas nas entrevistas



Fonte: Elaborada pelo autor (2020)

importante nesse sentido são os protocolos de transferências de animais entre viveiros, que notificam os carcinicultores quando o limite de aclimação é excedido.

9.7 Correlação dos Dados

A fim de entender a correlação dos atributos do *dataset*, foi aplicado o coeficiente de correlação de *Spearman* (Equação 9.4), que busca a dependência estatística de duas variáveis (ZAR, 2005).

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (9.4)$$

onde $d_i = \text{rg}(X_i) - \text{rg}(Y_i)$ é a diferença entre os dois postos de cada observação e nn é o número de observações. r_s pode ter um valor entre +1 e -1 onde: Um valor de +1 significa uma associação de classificação perfeita; Valor 0 significa que não há associação de classificação; Um valor -1 significa uma associação negativa perfeita entre os intervalos.

A partir dos resultados observou-se que *dt* (dureza total) está 71% relacionada com a ração posta no viveiro. Isso acontece porque a ração é misturada a um componente que ajuda na mudança de estado da água. Já a quantidade de ração inserida no viveiro está apenas 80% relacionada ao peso médio do camarão. Isso acontece pois há períodos

que o camarão come mais que outros períodos. Correlações abaixo de 70% foram desconsideradas, pois não são significativas para o estudo. Também foram desconsideradas correlações como dia de cultivo e ração inserida, pois consequentemente o arraçoador aumenta a quantidade ração com o passar dos dias.

9.8 Modelos de Predição

Por meio das entrevistas com produtores foi possível levantar os principais requisitos de BI, a partir dos quais foi possível criar modelos de predição voltadas para ajudar no planejamento da produção nas fazendas. Apesar de existirem outros requisitos, essa pesquisa limitou-se aos dois mais importantes, descritos a seguir. Cada requisito foi convertido em uma tarefa ou problema de predição. Para tanto, foi necessário treinar um algoritmo para cada tarefa. Para o treinamento, foram usados todos os atributos do *dataset*, exceto a data.

9.8.1 *Peso médio dos animais no tanque*

Essa tarefa pretende prever o peso médio dos animais no tanque no momento atual, independentemente da data da última biometria. Com essa informação é possível prever e planejar uma data para a despesca dos tanques, bem como mensurar a produção de cada um. Dado que peso é uma variável numérica, a técnicas de regressão foi escolhida.

9.8.2 *Quantidade de ração necessária para cada tanque*

Essa tarefa pretende prever a quantidade de ração necessária para alimentar os animais no tanque, pois, como visto na correlação, o consumo pode variar dependendo da época, idade, tamanho do animal, entre outros fatores. Essa informação pode trazer muita economia, pois evita desperdícios. A ração é um dos principais fatores de custo na produção do crustáceo. Esse também é um problema de regressão, pois busca uma variável numérica.

9.9 Algoritmos Utilizados

A fim de avaliar os melhores resultados e selecionar o melhor algoritmo para cada caso, foram testados os seguintes algoritmos. Como ambas as tarefas tratam-se de

problemas de regressão, os algoritmos usados foram os mesmos.

Regressão Linear é um modelo de predição estatística que trata dados dispersos de forma linear. O algoritmo gera uma linha de melhor ajuste no gráfico de dispersão, que representa a relação das variáveis explicativas com a variável dependente. A partir da linha de melhor ajuste é possível prever um valor para uma variável dependente dada uma nova entrada (LONG *et al.*, 1993). A Equação 9.5 representa a linha de melhor ajuste.

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (9.5)$$

Onde Y_i é a variável dependente; α é uma constante que representa a interceptação da linha com o eixo vertical; β é outra constante que representa a inclinação da reta (coeficiente angular); X_i é a variável explicativa (independente) que representa o quão essa variável explica o valor da variável dependente; ε_i é uma variável aleatória que representa os possíveis erros.

Árvore de Decisão (do inglês Decision Tree) é um modelo baseado no ganho de informação, que é calculado através de técnicas de categorização como Gini, Qui-quadrado ou Entropia. O ganho é expressado pela Equação 9.6.

$$G(A) = I(p, n) \sum_v \frac{p_i + n_i}{p + n} I(p_i; n_i) \quad (9.6)$$

onde:

$$I(p, n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \quad (9.7)$$

sendo p e n instâncias positivas e negativas de uma variável de classificação dicotômica (LONG *et al.*, 1993).

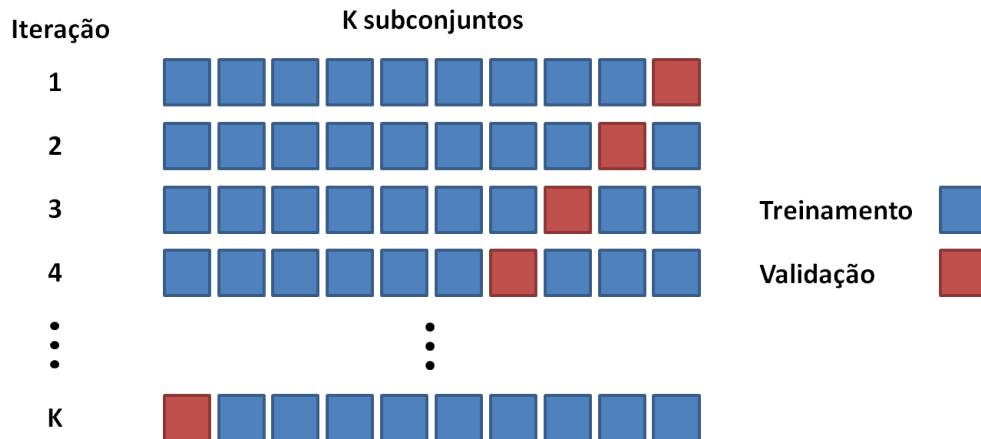
Floresta Aleatório (do inglês, Random Forest) é um algoritmo de aprendizagem conjunta que gera uma floresta de sub-árvores menores usando diversas técnicas de categorização, como Gini, Qui-quadrado e Entropia. O modelo combina as sub-árvores geradas usando a média dos resultados como saída (J.HAM; KAMBER, 2011).

9.9.1 Treinamento e Testes

Para testar e comparar os algoritmos, foi usado um método de treino/teste baseado em validação cruzada. O método divide o *dataset* em k partes. São realizadas k iterações

de treinamento e teste. Cada iteração separa a k -ésima parte para testar o modelo treinado. Depois, a parte de treino (k_i) é devolvida ao *dataset* enquanto a próxima parte (k_{i+1}) é separada para treino, sucessivamente, como poder ser observado na Figura 18.

Figura 18 – Ilustração Genérica da Validação Cruzada



Fonte: Extraída do site registro.br (2020)

Assim, ao final do processo, todos os exemplos são treinados e testados. Para esse trabalho, cada algoritmo foi testado por meio de uma validação cruzada com $k = 10$. Apesar de existirem outras formas de validação, onde algumas delas separa o *dataset* em treinamento e teste, optou-se pela validação cruzada por conta da limitação na quantidade de dados (BROWNE, 2000).

9.9.2 Métricas de Comparação

A fim de definir o melhor modelo para o contexto do trabalho, foram consideradas as seguintes métricas, amplamente usadas na literatura, para calcular a eficácia dos algoritmos:

- **Rsquare (R^2)** corresponde a uma medida de ajuste de um modelo estatístico linear generalizado em relação aos valores observados, variando de 0 a 1. Também conhecido como coeficiente de determinação. Esta métrica representa quanto o modelo consegue explicar os valores observados. Assim, quanto maior for esta métrica, mais explicativo é o modelo e melhor ele se ajusta à amostra (KASSAMBARA, 2018);
- **Erro Absoluto Médio (Mean Absolute Error - MAE)** mede a magnitude médias dos erros em um conjunto de previsões. Conforme mostrado na Equação 9.8, esta métrica se refere à média sobre a amostra de teste das diferenças absolutas entre a previsão (y_j) e a observação real ($\hat{y}_j$), em que todas as diferenças individuais têm

peso igual (KASSAMBARA, 2018);

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (9.8)$$

- **Erro Quadrático Médio da Raiz (Root-Mean Squared Error - RMSE)** é uma regra de pontuação quadrática que também mede a magnitude média do erro e representa a raiz quadrada da média das diferenças quadradas entre a previsão e observação real (Equação 9.9).

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (9.9)$$

Conforme mostrado na Equação 9.9, é importante destacar que a raiz quadrada dos erros quadráticos médios tem algumas implicações interessantes, já que os erros são elevados antes da média, atribuindo um peso relativamente alto a erros grandes. Neste sentido, isso justifica a utilidade desta métrica na abordagem deste trabalho, uma vez que erros grandes são particularmente indesejáveis.

Tanto o *MAE* quanto o *RMSE* expressam o erro médio de previsão de modelo. Ambas as métricas podem variar de 0 a ∞ e são indiferentes em relação à direção dos erros. Além disso, elas possuem pontuações orientadas negativamente. Então, já que representam taxas de erro, valores mais baixos correspondem a modelos que erram menos e, portanto, são melhores.

9.10 Resultados dos Testes

Como tratam-se de dois modelos de predição distintas, os resultados foram separados em duas etapas. Para facilitar o entendimento, os nomes das tarefas foram simplificados. Onde lia-se "Peso médio dos animais no tanque", lê-se "Peso Médio Estimado". Já onde lia-se "Quantidade de ração necessária para cada tanque", lê-se "Ração Necessária".

9.10.1 *Peso Médio Estimado*

Como observado, o peso médio do animal é uma informação muito importante para tomada de decisão, principalmente porque o preço por quilo de camarão varia dependendo do seu tamanho/peso. Em alguns casos esse valor é linear, aumentando de preço a cada

grama. Por conta disso, esse trabalho desenvolveu uma ferramenta de previsão de peso médio estimado dos animais no viveiro a fim de auxiliar o gestor a programar suas despesas no melhor momento comercial, ajudando a obter mais lucro. A Tabela 3 mostra a comparação do desempenho dos algoritmos testados em relação a cada uma dessas métricas.

Tabela 3 – Resultados dos testes por algoritmo (Peso Médio Estimado)

Algoritmo	R ²	MAE	RMSE
Regressão Linear	0.57	0.62	1.84
Árvore de Decisão	0.88	0.37	0.74
Floresta Aleatória	0.94	0.31	0.55

Fonte: elaborado pelo autor (2020).

Como mostrado na Tabela 3, os resultados da regressão linear foi o pior em todas as métricas. Isso porque, como mostrado no gráfico da Figura 16, o peso médio dos animais não obedece uma linearidade tão expressiva. Já os algoritmos baseado no ganho de informação mostraram resultados melhores, principalmente a Floresta Aleatória, pois conseguiu explicar 94% das variáveis explanatórias. Por isso, apesar de ter um custo computacional maior, o modelo de floresta aleatória foi escolhido para prever o peso médio estimado no sistema.

9.10.2 Ração Necessária

Como a ração é um dos itens de maior custo na produção do crustáceo, os produtores evitam seu desperdício. Contudo, diversos fatores podem influenciar no consumo da ração pelos animais, que vão desde fisiológicos, como troca de carapaça, a astrológicos, como a gravidade lunar, conhecido como "força da lua". Para auxiliar os técnicos responsáveis pela alimentação dos viveiros, foi desenvolvida um modelo para prever a quantidade de ração necessária em cada viveiro.

Tabela 4 – Resultados dos testes por algoritmo (Ração Necessária)

Algoritmo	R ²	MAE	RMSE
Regressão Linear	0,57	1.96	143.95
Árvore de Decisão	0,98	0.40	12.46
Floresta Aleatória	0,98	0.35	10.03

Fonte: elaborado pelo autor (2020).

Novamente o algoritmo que se destacou foi a Floresta Aleatória. Os algoritmos baseados em ganho de informações apresentaram uma vantagem muito significativa devido à complexidade do caso. No entanto, apesar de explicar cerca de 98% das variáveis explicativas, todos os algoritmos apresentam erros médios muito significativos, prejudicando o resultado da predição. Embora os algoritmos errarem poucas vezes, quando erram, acabam errando muito. Contudo, os resultados ainda são promissores.

10 EVOLUÇÃO DA PLATAFORMA

Desde a concepção, a plataforma já passou por várias etapas. O entendimento do manejo, mercado e metodologias da área de carcinicultura permitiram avançar com a solução, expandindo seus módulos. A Figura 19 mostra a evolução da plataforma com a liberação de novos módulos.

Figura 19 – Histórico de Lançamentos da Plataforma Crevettic



Fonte: Elaborado pelo autor (2020).

A primeira versão beta (MVP 0.1) da plataforma foi disponibilizada em janeiro de 2020 apenas com o módulo de ocorrências. O módulo permitia cadastrar usuário, propriedade (fazenda) e ocorrências, que podem caracterizar problemas na produção ou manejo do camarão, incluindo prejuízos. O módulo foi construído em parceria com o PSF Camarão a fim de coletar informações sobre a situação de saúde dos animais nas fazendas da região. Posteriormente, o módulo foi atualizado conforme novas demandas do projeto em março de 2020. A Figura 20 mostra interface da plataforma Crevettic.

Figura 20 – Telas da primeira versão da plataforma Crevettic

The figure shows three screenshots of the Crevettic platform interface:

- Left Screenshot (User Registration):** A form with fields for Username, Email, Endereço de e-mail, Senha, and Senha (confirm). A blue 'Cadastrar' button is at the bottom, along with a 'Fazer Login' link.
- Middle Screenshot (Property Registration):** A form with fields for Nome da Fazenda, Produtora de Camarão, Responsável, Descrição, Cooperativa, Endereço, Telefone, and CNPJ. A map of Ceará, Brazil, is shown at the bottom.
- Right Screenshot (Occurrence List):** A table titled 'Listar Ocorrências' with a 'Nova Ocorrência' button. The table has columns: Ano, Mês, Grau do Problema, Unidades Atingidas, and Descrição.

Ano	Mês	Grau do Problema	Unidades Atingidas	Descrição
2019	12		1	teste
2019	10		1	Ocorreu mortalidade alta
2019	11	2	11	baixo crescimento

Fonte: Elaborado pelo autor (2020).

A segunda versão beta (MVP 0.2) da plataforma foi lançada em abril de 2020, incluindo o módulo de qualidade da água, funcionalidade mais usada pelas fazendas segundo pesquisa realizada no Capítulo 6. Também foi desenvolvida uma tela de relatório com um gráfico multiparâmetros que permite acompanhar a variação da qualidade da água. A princípio, o módulo foi disponibilizado apenas para três fazendas a fim de coletar *feedbacks* e sugestões. A partir do uso foram relatados alguns erros, que foram resolvidos nas versões intermediárias (0.2.1 a 0.2.8).

Posteriormente foi liberada a versão beta (MVP 0.3) no mês de julho de 2020 com o módulo de biometria, permitindo aos carcinicultores realizarem o acompanhamento do crescimento do camarão. Logo depois foi agregado o módulo de controle de insumos na versão beta (MVP 0.4) liberado no mês de setembro do mesmo ano. Depois da liberação dos módulos também foram relatados diversos erros e sugestões de melhorias, como a tela de *dashboard* e um novo gráfico na tela de relatório. Esses incrementos foram incluídos em versões intermediárias. Observou-se que durante as medições dos animais também eram feitas algumas análises de biopatometria, que consistiam em verificar o estado físico do crustáceo a olho nu. Então, esses parâmetros foram incluídos ao módulo de biometria (MVP 0.3.12).

Assim, com esses módulos desenvolvidos, o sistema já tratava os principais dados técnicos gerados pelas fazendas de camarão. A partir desse ponto, o foco do trabalho foi melhorar os módulos existentes, expandindo o teste para mais fazendas a fim de validar melhor a plataforma e melhorar seu uso.

11 CONCLUSÕES E TRABALHOS FUTUROS

A carcinicultura é uma área que envolve muitas variáveis, tornando as decisões difíceis. Como pode ser visto nesse trabalho, boa parte das informações são aproximações estatísticas do valor real. Portanto, o uso da computação é essencial no gerenciamento e processamento de dados importantes para tomada de decisão.

A pesquisa mostrou que o uso de técnicas de mineração de dados pode agregar valor significativo à tomada de decisão nas fazendas produtoras de camarão. Os resultados apresentados pelos modelos preditivos são um exemplo disso, apesar de não predizerem perfeitamente. Contudo, a limitação dos dados usados para o treinamento influenciaram na qualidade dos modelos. Espera-se que, com mais dados levantados, os modelos apresentem melhores resultados.

O desafio de construir uma ferramenta realmente útil para a comunidade de carcinicultura foi um dos grandes desafios desse trabalho, além do desafio tecnológico em si. Para tanto buscou-se conhecimento em diversas áreas, incluindo negócios, gestão e a própria carcinicultura. As parcerias construídas além dos muros da universidade também foram cruciais para o sucesso do projeto, pois mostraram a realidade do setor.

A Plataforma Crevettic, desenvolvida a partir dessa pesquisa, se tornou uma *startup* de *agrotech* com a missão de melhorar a qualidade e produtividade das fazendas de camarão nacionais, ainda carentes de tecnologia. Futuramente a plataforma deve expandir seu produto para outras áreas da aquicultura, como a piscicultura. Já foram realizadas algumas entrevistas com produtores e especialistas da área, visando estabelecer novas parcerias nesse sentido.

Também já se trabalha na melhoria do Módulo de BI, implementando novos gráficos e relatórios. Além disso, também estão sendo concebidos novos modelos de predição, bem como a melhoria dos existentes. Para tanto, pretende-se enriquecer o *dataset* com novos dados, tais como geográficos, climáticos, entre outros. Além disso, pretende-se desenvolver um sistema de consultoria inteligente capaz de dar recomendações corretivas baseadas nas análises preventivas levantadas por consultores.

12 PRODUTIVIDADE ACADÊMICA E CONQUISTAS

Durante o mestrado pôde-se desenvolver diversos trabalhos acadêmicos, bem como participações em eventos de divulgação científica, tecnológica e de inovação. Abaixo seguem alguns exemplos.

12.1 Trabalhos publicados

- Analyzing Patterns of a Bicycle Sharing System for Generating Rental Flow Predictive Models, COURB - SBRC (VIANA *et al.*, 2019);
- Smart "Health of Things": a Model Based in Data Mining for an IoT Health System Used in Hospital and Home Urgencies, WebMedia (BRAGA *et al.*, 2019b);
- Sistema Baseado em Heurística para Escolha de Melhor Entregador em Rotas Dinâmicas no Contexto de Entrega de Comida, SBPO (BRAGA *et al.*, 2019a);
- An Architecture's Propose Based on Linked Data and IoT to Improve the Health Service, EATIS (FREITAS *et al.*, 2020);
- Mobile Application Based on Artificial Intelligence in the Context of Covid 19, CONINP (ALVES *et al.*, 2020).

12.2 Trabalhos apresentados

- Analyzing Patterns of a Bicycle Sharing System for Generating Rental Flow Predictive Models, COURB - SBRC (VIANA *et al.*, 2019);
- Smart "Health of Things": a Model Based in Data Mining for an IoT Health System Used in Hospital and Home Urgencies, WebMedia (BRAGA *et al.*, 2019b).

12.3 Capítulo de livros

- Crevettic: Uma Plataforma de Business Intelligence para Gestão de Carcinicultura, SEMAQUI 2020.
- Icapuí Conectado ao Novo Mundo, PLD Icapuí 2021 (ainda não publicado)

12.4 Participações em palestras e *lives*

- Crevettic: Uma Plataforma de Business Intelligence para Gestão da Carcinicultura, IX SEMAP.
- Como criar registros de distribuição espaço-temporal de problemas na carcinicultura?
- Como gerenciar os dados da carcinicultura com o uso da informática?
- Como funciona a ferramenta digital de acompanhamento espaço-temporal das ocorrências dos criadores de camarão?

12.5 Conquistas

Além dos trabalhos científicos, esse projeto passou por diversos programas e participou de vários concursos.

- Centelha é um programa promovido pelo Ministério da Ciência, Tecnologia e Inovações (MCTI) e pela Financiadora de Estudos e Projetos (Finep) com o objetivo de incentivar a abertura de novas empresas. A plataforma Crevettic foi submetida no Centelha - Ceará, concorrendo com quase mil outras propostas, sendo contemplada pelo programa com um aporte de R\$ 80 mil.
- Startup CE é um programa de aceleração promovido pelo Sebrae Ceará em parceria com outras entidades. A Plataforma Crevettic ficou entre as 25 ideias selecionadas para segunda etapa do programa entre mais de 100 startups que concorreram na primeira fase.
- Corredores Digitais é um programa integrado de aceleração realizado pela Secretaria da Ciência, Tecnologia e Educação Superior (SECITECE) voltado para estudantes que objetivam transformar suas ideias ou pesquisas em negócios. A Plataforma Crevettic foi selecionada para etapa de "tração de negócio".
- A Plataforma Crevettic também foi selecionada como prova de conceito pelo projeto Revive Negócio, mantido pela Secretaria do Desenvolvimento Econômico e Trabalho (SEDET) focado no desenvolvimento de novos negócios, onde obteve aconselhamento e consultorias.
- A proposta da plataforma Crevettic foi contemplada pelo Edital de Apoio a Inovação número 38/2019 realizado pela Pró-Reitoria de Pesquisa e Pós-Graduação da Universidade Federal Rural do Semi-Árido.

REFERÊNCIAS

- ALVES, R.; BRAGA, O.; FONSÊCA, O.; MOREIRA, M.; RODRIGUES, J.; SILVEIRA, R.; OLIVEIRA, M.; NETO, A. Mobile application based on artificial intelligence in the context of covid 19. In: IFCE. **CONGRESSO INTERNACIONAL VIRTUAL DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO DO INSTITUTO FEDERAL DO CEARÁ**. [S.l.], 2020.
- ARAUJO, J. A.; NORÕES, A. K. M.; MONTEIRO, J. V.; ARAÚJO, R. C. P. de; SILVA, F. P. da. Eficiência Produtiva das Fazendas de Carcinicultura no Estado do Ceará. **Revista de Economia e Sociologia Rural**, SciELO Brasil, v. 56, n. 1, p. 35–50, 2018.
- AWAD, M.; KHANNA, R. **Efficient learning machines: theories, concepts, and applications for engineers and system designers**. [S.l.]: Springer Nature, 2015.
- AWAD, M.; KHANNA, R. Machine learning and knowledge discovery. In: **Efficient Learning Machines**. [S.l.]: Springer, 2015. p. 19–38.
- BOONE, L. **Anomuran, macruran Crustacea from Panama and Canal Zone**. [S.l.]: American Museum of Natural History, 1931.
- BRAGA, O.; ALBUQUERQUE, G.; OLIVEIRA, M.; MONTEIRO, O. Intelligent solution for classification of diseases transmitted by vector aedes aegypti. In: **Proceedings of the Euro American Conference on Telematics and Information Systems**. New York, NY, USA: ACM, 2018. (EATIS '18), p. 8:1–8:5. ISBN 978-1-4503-6572-7. Disponível em: <<http://doi.acm.org/10.1145/3293614.3293640>>.
- BRAGA, O. C.; MOREIRA, C. R.; MACEDO, F. de; JÚNIOR, F. C. de L. Sistema baseado em heurística para escolha de melhor entregador em rotas dinâmicas no contexto de entrega de comida. In: UNICAMP. **LI SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL**. [S.l.], 2019.
- BRAGA, O. C.; NETO, F. M. M.; OLIVEIRA, A. M. B. de; FILHO, R. V. C. Smart"health of things"a model based in data mining for an iot health system used in hospital and home urgencies. In: **Proceedings of the 25th Brazillian Symposium on Multimedia and the Web**. [S.l.: s.n.], 2019. p. 89–92.
- BROWNE, M. W. Cross-validation methods. **Journal of mathematical psychology**, Elsevier, v. 44, n. 1, p. 108–132, 2000.
- BURSTEIN, F.; HOLSAPPLE, C. W. **Handbook on decision support systems 2: variations**. [S.l.]: Springer Science & Business Media, 2008.
- CEARÁ, L. N. d. E. do. Censo da carcinicultura do litoral norte do estado do ceará e zonas interioranas adjacentes. 2017.
- CECI, F. Business intelligence. **Palhoça, Unisul**, 2012.
- COSTA, S. W. da; VICENTE, L. R. M.; SOUZA, T. M. de; ANDREATTA, E. R.; MARQUES, M. R. F. Parâmetros de cultivo e a enfermidade da mancha-branca em fazendas de camarões de santa catarina. **Pesquisa Agropecuária Brasileira**, v. 45, n. 12, p. 1521–1530, 2011.

- DUY, N. T. K.; TU, N. D.; SON, T. H.; KHANH, L. H. D. Automated monitoring and control system for shrimp farms based on embedded system and wireless sensor network. In: IEEE. **2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)**. [S.l.], 2015. p. 1–5.
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. de. **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. Rio de Janeiro, RJ, Brasil: LTC, 2015.
- FAO, F. Food and agriculture organisation of the united nations. **Retrieved on**, v. 15, 2012.
- FAYYAD, U. M.; PIATETSKY-SHAPIO, G.; SMYTH, P.; UTHURUSAMY, R. *et al.* **Advances in knowledge discovery and data mining**. [S.l.]: AAAI press Menlo Park, 1996. v. 21.
- FREITAS, R.; BRAGA, O.; ANDRADE, O.; MENDES, M. O. M.; BARRETO, I. An architecture's propose based on linked data and iot to improve the health service. **EATIS 2020 10th Euro American Conference on Telematics and Information Systems**, ACM Digital Library, v. 10, 2020.
- HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**. [S.l.]: Elsevier, 2011.
- IBGE. **Produção da Pecuária Municipal 2018**. 2019. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/84/ppm_2018_v46_br_informativo.pdf>. Acesso em: 10 março. 2020.
- J.HAM; KAMBER, M. **Data Mining: Concepts and Techniques**. 3. ed. [S.l.]: Morgan Kaufmann Publishers is an imprint of Elsevier, 2011.
- KASSAMBARA, A. **Machine Learning Essentials: Practical Guide in R**. [S.l.]: STHDA, 2018.
- KNAPP, J.; ZERATSKY, J.; KOWITZ, B. **Sprint: O método usado no Google para testar e aplicar novas ideias em apenas cinco dias**. [S.l.]: Editora Intrínseca, 2017.
- LIGHTNER, D.; LIGHTNER, D.; LIGHTNER, D.; LIGHTNER, D.; LIGHTNER, D. A handbook of shrimp pathology and diagnostic procedures for diseases of cultured penaeid shrimp. 1996.
- LONG, W. J.; GRIFFITH, J. L.; SELKER, H. P.; D'AGOSTINO, R. B. A comparison of logistic regression to decision-tree induction in a medical domain. **Computers and Biomedical Research**, Elsevier Science, v. 26, n. 1, p. 74–97, 1993.
- MORALES, V.; CUÉLLAR, A. Guía técnica–patología e inmunología de camarones penaeidos. eds. **Panamá (República de Panamá)**, p. 237–308, 2014.
- NEVES, S. R. d. A. **O programa de saúde nas fazendas de camarão PSF_CAMARO e os seus impactos sobre os desempenhos produtivos e econômicos na carcinicultura familiar do baixo rio Pirangi, Ceará, Brasil**. Tese (Doutorado) — Universidade Federal do Ceará, 2018.
- PIPLANI, D.; SINGH, D. K.; SRINIVASAN, K.; RAMESH, N.; KUMAR, A. *et al.* Digital platform for data driven aquaculture farm management. In: ACM. **Proceedings of the 7th International Conference on HCI, IndiaHCI 2015**. [S.l.], 2015. p. 95–101.

REZENDE, F. P.; MATAVELI, M. Impactos da mancha branca nos custos de produção do camarão no nordeste. **Embrapa Pesca e Aquicultura-Outras publicações técnicas (INFOTECA-E)**, Brasília, DF: CNA, 2017., 2017.

RIES, E. **A startup enxuta**. [S.l.]: Leya, 2012.

SANTOS, M. Y.; RAMOS, I. **Business Intelligence: tecnologias da informação na gestão de conhecimento**. [S.l.]: FCA-Editora de Informática, Lda, 2006.

SENAR. **Larvicultura de camarão marinho: do náuplio a pós-larva**. 2016. Disponível em: <<https://www.cnabrazil.org.br/assets/arquivos/166-LARVICULTURA.pdf>>. Acesso em: 28 novembro. 2020.

SHAREEF, Z.; REDDY, S. Wireless sensor network for aquaculture: Review, survey, and case study of aquaculture practices in western godavari region. **Journal of Ambient Intelligence and Smart Environments**, v. 10, p. 409–423, 09 2018.

SOSA, B.; FERREIRA, P. B.; HOSHIBA, M. A.; WISCHRAL, D. Parâmetros físico-químicos da água mais relevantes na produtividade da criação de camarões em bioflocos. **Anais do Salão Internacional de Ensino, Pesquisa e Extensão**, v. 7, n. 2, 2016.

STANGE, R.; JOSÉ, J. Reconhecimento de padrões em classificadores—comparação de técnicas e aplicações. In: **IV Workshop de Tecnologia Adaptativa (WTA 2010), January 21-22, Sao Paulo, SP, Brazil**. [S.l.: s.n.], 2010. p. 63–67.

SUTHERLAND, J. **Scrum: a arte de fazer o dobro do trabalho na metade do tempo**. [S.l.]: Leya, 2016.

TAN, P.; STEINBACH, M.; KUMAR, V.; FERNANDES, A. **Introdução ao datamining: mineração de dados**. Ciencia Moderna, 2009. ISBN 9788573937619. Disponível em: <<https://books.google.com.br/books?id=69d6PgAACAAM>>.

TURBAN, E.; SHARDA, R.; ARONSON, J. E.; KING, D. **Business intelligence: um enfoque gerencial para a inteligência do negócio**. [S.l.]: Bookman Editora, 2009.

VIANA, J. D. F.; BRAGA, O. C.; CHAVES, L.; NETO, F. M. M. *et al.* Analyzing patterns of a bicycle sharing system for generating rental flow predictive models. In: SBC. **Anais do III Workshop de Computação Urbana**. [S.l.], 2019. p. 57–70.

XI, M.; ADCOCK, M.; MCCULLOCH, J. An end-to-end augmented reality solution to support aquaculture farmers with data collection, storage, and analysis. In: ACM. **The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry**. [S.l.], 2019. p. 37.

ZAR, J. H. Spearman rank correlation. **Encyclopedia of Biostatistics**, Wiley Online Library, v. 7, 2005.