

UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Computação



Dissertação

**Utilizando um Classificador de Sentimento Baseado em Léxico em Negociações
de Jogos de RPG para Fazer Predições no Contexto de Recursos Naturais**

Karine Pestana Ramos

Pelotas, 2021

Karine Pestana Ramos

**Utilizando um Classificador de Sentimento Baseado em Léxico em Negociações
de Jogos de RPG para Fazer Predições no Contexto de Recursos Naturais**

Dissertação apresentada ao Programa de Pós-Graduação em Computação do Centro de Desenvolvimento Tecnológico da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Marilton Sanchotene de Aguiar
Coorientadora: Profa. Dra. Larissa Astrogildo de Freitas

Pelotas, 2021

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

R175u Ramos, Karine Pestana

Utilizando um classificador de sentimento baseado em léxico em negociações de jogos de RPG para fazer predições no contexto de recursos naturais / Karine Pestana Ramos ; Marilton Sanchotene de Aguiar, orientador ; Larissa Astrogildo de Freitas, coorientadora. — Pelotas, 2021.

63 f.

Dissertação (Mestrado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2021.

1. Aprendizado de máquina. 2. Análise de sentimentos. 3. Jogos de papéis. 4. Gerenciamento de recursos hídricos. I. Aguiar, Marilton Sanchotene de, orient. II. Freitas, Larissa Astrogildo de, coorient. III. Título.

CDD : 005

Karine Pestana Ramos

**Utilizando um Classificador de Sentimento Baseado em Léxico em Negociações
de Jogos de RPG para Fazer Predições no Contexto de Recursos Naturais**

Dissertação aprovada, como requisito parcial, para obtenção do grau de Mestre em Ciência da Computação, Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas.

Data da Defesa: 8 de abril de 2021

Banca Examinadora:

Prof. Dr. Marilton Sanchotene de Aguiar (orientador)

Doutor em Computação pela Universidade Federal do Rio Grande do Sul.

Profa. Dra. Larissa Astrogildo de Freitas (coorientadora)

Doutora em Ciência da Computação pela Pontifícia Universidade Católica do Rio Grande do Sul.

Prof. Dra. Diana Francisca Adamatti

Doutora em Engenharia Elétrica pela Escola Politécnica da Universidade de São Paulo.

Prof. Dr. Tiago Thompsen Primo

Doutor em Computação pela Universidade Federal do Rio Grande do Sul.

Quero dedicar este trabalho e todo o esforço envolvido em seu desenvolvimento para todas as pessoas que sofrem com transtorno de ansiedade e outras doenças mentais. Dedico aos que persistem em realizar o que desejam. Dedico aos que respeitam seu tempo e não desistem da vida. É uma vitória chegar até aqui.

AGRADECIMENTOS

São inúmeros agradecimentos. Primeiramente quero agradecer à minha família, meus pais e minha irmã. É um privilégio ter o apoio de vocês ao longo desse período. Em especial quero agradecer a minha mãe, que acompanhou de perto o processo de conclusão desta Dissertação e as dificuldades técnicas e pessoais envolvidas. Ainda no âmbito familiar quero agradecer à minha tia Sônia, que mesmo de longe sempre olhou por mim e apoiou para que fosse possível a conclusão dos meus estudos. Quero agradecer aos meus amigos cujo os momentos de descontração, desabafos e compartilhamentos permitiram que eu tivesse forças para concluir este trabalho. Também quero agradecer ao meu parceiro de vida e amor Juliano, que me deu apoio e suporte em momentos que eu me vi esgotada. Por último quero agradecer aos meus orientadores, Marilton e Larissa, que a cada reunião, encontro ou conversa sempre procuravam me animar e renovavam minhas forças para seguir com este trabalho. Agradeço por ter tido ambos como orientadores, não poderia ter orientadores mais pacientes e compreensivos. Agradeço incansavelmente a todos aqui citados, sem vocês não seria possível chegar até aqui.

*“Não é o fracasso que nos detém, mas é a relutância em
recomeçar que nos faz estagnar.”*

— CLARISSA PINKOLA ESTÉS

RESUMO

RAMOS, Karine Pestana. **Utilizando um Classificador de Sentimento Baseado em Léxico em Negociações de Jogos de RPG para Fazer Predições no Contexto de Recursos Naturais**. Orientador: Marilton Sanchotene de Aguiar. 2021. 63 f. Dissertação (Mestrado em Ciência da Computação) – Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2021.

O gerenciamento de recursos naturais concentra-se especificamente no entendimento científico e técnico destes recursos, sua ecologia e na capacidade de suporte à vida destes recursos. Nesse contexto, a simulação aparece como um mecanismo que reproduz sistemas reais usando ferramentas de computação como apoio à previsão de consequências futuras, auxiliando na tomada de decisões. Jogos de papéis são um tipo de jogo em que os jogadores interpretam personagens criados dentro de um ambiente, que organiza as ações dos jogadores, determinando os limites do que pode ou não ser feito. No âmbito dos recursos hídricos, o RPG auxilia no processo de entendimento do problema e como cada jogador pensa/analisa/observa o mesmo sob a sua perspectiva e assim elabora as estratégias que julga pertinente ao seu papel e ao jogo como um todo. Esta dissertação está inserida no contexto de um projeto financiado pela CAPES e ANA, onde um jogo de RPG está em desenvolvimento para a gestão participativa dos recursos hídricos. A análise de sentimentos refere-se ao uso de processamento de língua natural para identificar sistematicamente estados afetivos e informações subjetivas, a fim de criar conhecimento estruturado que possa ser utilizado por um sistema de apoio ou tomador de decisão. Constatou-se a inexistência de trabalhos que utilizam técnicas para análise de sentimentos para a adição de atributos subjetivos para predição de comportamentos em jogos, no caso para a predição dos níveis de poluição. O objetivo geral desta dissertação é a construção de modelos de predição dos níveis de poluição com o uso de atributos subjetivos, através da integração de técnicas de análise de sentimento para a classificação das transações realizadas pelos jogadores. Os modelos foram gerados utilizando-se das teorias das emoções propostas por Paul Ekman e Robert Plutchik. Através das métricas de avaliação utilizadas os modelos construídos demonstram bons resultados, permitindo que ao longo da partida do RPG, considerando os diálogos das negociações, os modelos possam prever os níveis de poluição, visto que, as transações somente são concluídas ao fim de cada rodada.

Palavras-chave: Aprendizado de máquina. Análise de sentimentos. Jogos de papéis. Gerenciamento de recursos hídricos.

ABSTRACT

RAMOS, Karine Pestana. **Using a Lexicon-Based Sentiment Classifier in RPG Game Negotiations to Make Predictions in the Context of Natural Resources.** Advisor: Marilton Sanchotene de Aguiar. 2021. 63 f. Dissertation (Masters in Computer Science) – Technology Development Center, Federal University of Pelotas, Pelotas, 2021.

The management of natural resources focuses specifically on the scientific and technical understanding of these resources, their ecology, and the ability to support these resources' lives. In this context, simulation appears as a mechanism that reproduces natural systems using computer tools to predict future consequences, to assist in decision making. Role-playing games are a type of game in which players interpret characters created within an environment, which organizes the players' actions, determining the limits of what they can do or not. In the scope of water resources, RPG helps understand the problem and how each player thinks, analyzes, and observes the same from his perspective and thus elaborates the strategies he deems relevant to his role and the game. We developed this dissertation in the context of a project financed by CAPES and ANA, where an RPG game is under development for the participatory management of water resources. The analysis of feelings refers to natural language processing to systematically identify affective states and subjective information to create structured knowledge that a support system or decision-maker can use. We did not find studies that use techniques for analyzing feelings to add subjective attributes to predict behavior in games, in this case, to indicate pollution levels. The general objective of this dissertation is to construct models for the prediction of pollution levels with the use of subjective attributes through the integration of sentiment analysis techniques for the classification of transactions carried out by players. We created models using the theories of emotions proposed by Paul Ekman and Robert Plutchik. Through the evaluation metrics used, the built models show promising results. Throughout the RPG match, the models can predict the pollution levels considering the negotiation dialogues at the end of each round.

Keywords: Machine learning. Sentiment analysis. Role-playing games. Water resource management.

LISTA DE FIGURAS

Figura 1	Diagrama de agentes no ambiente do jogo. Fonte: (LEITZKE et al., 2019).	19
Figura 2	Níveis de poluição do ambiente. Fonte: Manual do RPG GORIM ¹ . .	21
Figura 3	Diagrama das etapas realizadas para a construção do novo <i>dataset</i> . Fonte: Própria.	36
Figura 4	Exemplo de um diálogo original na primeira rodada de uma das partidas do GORIM. Fonte: Própria.	37
Figura 5	Exemplo de uma transação isolada das demais e padronizada. Fonte: Própria.	37
Figura 6	Exemplo da busca em níveis do termo Alegria. Fonte: Própria . . .	38
Figura 7	Exemplo da classificação por sentença – LS Ekman. Fonte: Própria	40
Figura 8	Exemplo da classificação por sentença – LS Plutchik Fonte: Própria	41
Figura 9	Diagrama das etapas realizadas para união dos <i>datasets</i> . Fonte: Própria	42
Figura 10	Diagrama das etapas realizadas para o desenvolvimento dos modelos de AM. Fonte: Própria.	43
Figura 11	Cenário 1: Amostra do Modelo de Regressão Linear. Fonte: Própria.	45
Figura 12	Cenário 1: Amostra do Modelo de <i>Decision Tree Regression</i> . Fonte: Própria.	46
Figura 13	Cenário 1: Amostra do Modelo de <i>Random Forest Regression</i> . Fonte: Própria.	46
Figura 14	Cenário 1: Amostra do Modelo de <i>Support Vector Regression</i> . Fonte: Própria.	46
Figura 15	Cenário 2: Amostra do Modelo de Regressão Linear. Fonte: Própria.	47
Figura 16	Cenário 2: Amostra do Modelo de <i>Decision Tree Regression</i> . Fonte: Própria.	48
Figura 17	Cenário 2: Amostra do Modelo de <i>Random Forest Regression</i> . Fonte: Própria.	48
Figura 18	Cenário 2: Amostra do Modelo de <i>Support Vector Regression</i> . Fonte: Própria.	48

LISTA DE TABELAS

Tabela 1	Comparação entre os trabalhos relacionados.	32
Tabela 2	Exemplo de dado gerado pelo motor RPG GORIM.	35
Tabela 3	Exemplo dos dados reorganizados.	35
Tabela 4	Comparação entre os trabalhos relacionados e a abordagem proposta.	44
Tabela 5	Cenário 1: Avaliação dos modelos.	45
Tabela 6	Cenário 2: Avaliação dos modelos.	47

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
ANA	Agência Nacional das Águas
AS	Análise de Sentimentos
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
DFS	<i>Depth-First Search</i>
DWS	<i>Degree-Words Set</i>
EWS	<i>Emotion-Words Set</i>
IA	Inteligência Artificial
LAS	<i>Location-Areas Set</i>
LS	Léxico de Sentimento
LIWC	<i>Linguistic Inquiry and Word Count</i>
MAE	Erro Médio Absoluto
MSE	Erro Médio Quadrado
PLN	Processamento de Língua Natural
R^2	<i>Coefficient of Determination</i>
RFE	<i>Recursive Feature Elimination</i>
RPG	<i>Role-Playing Game</i>
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>

SUMÁRIO

1	INTRODUÇÃO	14
2	REFERENCIAL TEÓRICO E TECNOLÓGICO	17
2.1	Gestão de Recursos Hídricos	17
2.2	Jogos de Papéis e Jogos Sérios	18
2.3	Aprendizado de Máquina	21
2.4	Análise de Sentimentos	26
2.5	Teorias de Emoções	29
2.6	Trabalhos Relacionados	29
3	ABORDAGEM PROPOSTA	34
3.1	Construção dos <i>Datasets</i>	34
3.1.1	Dados do motor RPG GORIM	34
3.1.2	Diálogos entre os personagens fazendo as transações	36
3.2	Léxicos de Sentimentos	37
3.3	Dicionário de Termos Intensificadores	39
3.4	Classificador de Sentimentos	39
3.5	Preditores	42
4	RESULTADOS	44
4.1	Cenário 1	44
4.2	Cenário 2	46
5	CONCLUSÕES	49
	REFERÊNCIAS	51
	APÊNDICE A DIÁLOGOS ORIGINAIS	58
	APÊNDICE B DIÁLOGOS PRÉ-PROCESSADOS	61
	APÊNDICE C DICIONÁRIO DE TERMOS INTENSIFICADORES	63

1 INTRODUÇÃO

O gerenciamento de recursos naturais é uma área que busca melhores formas de gerenciar terras, águas, plantas e animais, baseado na qualidade de vida das pessoas no presente e para gerações futuras. Estes temas deram origem à noção de desenvolvimento sustentável, um princípio que forma a base para o gerenciamento da terra e a governança ambiental. O gerenciamento de recursos naturais concentra-se especificamente no entendimento científico e técnico de recursos ecológicos e na capacidade de suporte à vida desses recursos (HOLZMAN, 2009).

Os problemas criados pela emissão excessiva de poluentes não são recentes e aumentaram significativamente devido ao rápido crescimento industrial. Uma alternativa é tomar medidas preventivas, por exemplo, conhecendo o risco da instalação de uma nova indústria em uma região. O gerenciamento de recursos naturais exige que o governo, os cidadãos e demais atores interajam com o apoio da computação. Nesse contexto, a simulação aparece como um mecanismo que reproduz sistemas reais usando ferramentas de computação como apoio à previsão de consequências futuras, auxiliando na tomada de decisões (ADAMATTI; AGUIAR, 2012).

O uso de ferramentas computacionais é uma alternativa para prever riscos. Dessa forma, essas ferramentas podem criar novas possibilidades para obter uma melhor qualidade de vida. A utilização de ferramentas computacionais utilizando técnicas de Inteligência Artificial (IA), pode ser uma solução, visto que elas têm a flexibilidade necessária para tratar a dinâmica existente na área de recursos naturais.

Jogos de papéis (em inglês, *Role-playing games* – RPGs) são um tipo de jogo em que os jogadores interpretam personagens. Esses personagens são criados dentro de um cenário específico (um ambiente), que organiza as ações dos jogadores, determinando os limites do que pode ou não ser feito (BANDINI; MANZONI; VIZZARI, 2002). Dessa forma, os RPGs são jogos em que cada jogador desempenha um papel e toma decisões para atingir seus objetivos. De fato, os jogadores usam o RPG como um “laboratório social”, porque podem tentar várias possibilidades, sem consequências reais (BARRETEAU; PAGE; D’AQUINO, 2003).

No âmbito dos recursos hídricos, o RPG auxilia no processo de entendimento do

problema e como cada jogador pensa/analisa/observa o mesmo sob a sua perspectiva e assim elabora as estratégias que julga pertinente ao seu papel e ao jogo como um todo. Os personagens de um RPG respeitam um sistema de regras, que serve para organizar suas ações, determinando os limites do que pode ou não ser feito.

Buscando compreender comportamentos dos jogadores nos jogos, é importante em um primeiro momento compreender suas emoções. O processamento emocional é importante também para entender comportamentos relacionados à polaridade. A oportunidade de capturar automaticamente os sentimentos do público em geral sobre eventos sociais, movimentos políticos, campanhas de marketing e preferências de produtos aumentou o interesse da comunidade científica – devido às questões em aberto – e do mundo dos negócios pelas emergentes aplicações na previsão de mercados (CAMBRIA, 2016). As emoções desempenham um papel essencial na interação entre os personagens e os eventos que os envolvem (KIM; KLINGER, 2018a).

Liu (2010) define que a informação textual de um modo genérico pode ser categorizada em dois grandes tipos: fatos e opiniões. Fatos referem-se a expressões objetivas sobre entidades e suas propriedades. Opiniões são expressões subjetivas utilizadas para descrever os sentimentos das pessoas, através destas opiniões é possível expressar avaliações sobre entidades, eventos e suas propriedades. A ideia de opinião trata-se de um conceito muito amplo.

A análise de sentimentos, de acordo com Liu (2010) trata-se do estudo computacional de opiniões, sentimentos e emoções expressos em um texto. A análise de sentimentos engloba ainda a aplicação de técnicas computacionais de análise de texto e de Processamento de Língua Natural (PLN), utilizadas como forma de automatizar o processo de identificação e extração de informações subjetivas em um dado texto. Esta análise é aplicada em diversas áreas, como informações de consumidor, marketing, livros, aplicativos e sites (HUSSEIN, 2018). Por sua grande variedade de aplicações, a análise de sentimentos se torna uma técnica destaque para tomada de decisão.

No grupo de pesquisa onde este trabalho se insere, tem-se um Projeto financiado pelo Programa de Apoio ao Ensino e à Pesquisa Científica e Tecnológica em Regulação e Gestão de Recursos Hídricos – Pró-Recursos Hídricos Chamada 16/2017 da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Agência Nacional das Águas (ANA).

Neste Projeto está sendo desenvolvido um jogo de RPG para a gestão participativa dos recursos hídricos, utilizando como base dados do estado do Rio Grande do Sul, cuja aplicação-piloto se dá no Comitê de Gerenciamento das Bacias Hidrográficas da Lagoa Mirim e do Canal São Gonçalo, envolvendo as cidades de Rio Grande e Pelotas no Sul do Brasil (BORN, 2018; FARIAS et al., 2019, 2020; LEITZKE et al., 2019).

O RPG de mesa desenvolvido serve como base para coleta de dados desta dissertação. Neste jogo, as negociações entre os jogadores são chamadas de transações.

As transações possuem registros padronizados e não padronizados. Os registros padronizados referem-se ao leque de ações permitidas a cada agente do RPG. Assim cada ação pode ser formalmente registrada. Os registros não-padronizados referem-se a como os agentes negociam as ações que serão realizadas de forma padronizada. Tratando-se de um jogo de RPG para a gestão participativa dos recursos hídricos, os níveis de poluição do mundo (do ambiente do jogo) são utilizados para se medir quanto saudável está o meio ambiente. Altos níveis de poluição acarretam em perdas proporcionalmente significativas de produção para agricultores e empresários.

Esta dissertação tem como objetivo geral utilizar técnicas de análise de sentimentos na identificação de emoções considerando as teorias propostas por Paul Ekman (EKMAN, 1993) e Robert Plutchik (PLUTCHIK, 1991) nas negociações ocorridas entre os jogadores do RPG GORIM e incluir estas informações subjetivas na geração de modelos preditores dos níveis de poluição.

Este trabalho está organizado da seguinte maneira: no Capítulo 2 são apresentados o referencial teórico e tecnológico utilizados, incluindo os trabalhos relacionados sobre os conteúdos abordados neste estudo; no Capítulo 3 são apresentadas descrições das bases de dados utilizadas e das tecnologias e metodologias utilizadas; no Capítulo 4 são apresentados os resultados de acordo com os cenários propostos; e, por fim, no Capítulo 5 são apresentadas as considerações finais e os trabalhos futuros.

2 REFERENCIAL TEÓRICO E TECNOLÓGICO

Neste Capítulo são apresentados o referencial teórico e tecnológico utilizados neste trabalho, ou seja, os principais conceitos e os trabalhos relacionados a esta dissertação

2.1 Gestão de Recursos Hídricos

Partindo da premissa que é um direito de todos possuir acesso aos recursos hídricos, o gerenciamento dos recursos hídricos torna-se um dos grandes desafios que devem ser administrados pela sociedade na atualidade. Recursos hídricos são um dos recursos mais essenciais para o ecossistema e populações, é necessário para todos os tipos de atividades sociais e econômicas, para a vida e a saúde da humanidade (PONTE et al., 2016).

Por ser um bem público, os recursos hídricos servem para atender às necessidades essenciais da sobrevivência humana, cujo uso deve ser feito racionalmente, assegurando o desenvolvimento e a preservação do meio ambiente (PORTAL EDUCAÇÃO, 2020). No contexto de recursos hídricos, a bacia hidrográfica é a unidade básica físico-territorial de planejamento e gerenciamento dos recursos hídricos, sendo necessárias ações em uma administração sustentável de recursos hídricos que visem assegurar padrões de qualidade e quantidade dentro desta unidade. Apesar do Brasil possuir em seu território grandes bacias hidrográficas ainda existem conflitos acerca de distribuição e compartilhamento (FARIAS et al., 2019).

A gestão de recursos hídricos consiste em um conjunto de ações destinadas a regular o uso, o controle e a proteção dos recursos hídricos, considerando a legislação e normas pertinentes da sociedade. A gestão tem por objetivo a integração de projetos e atividades com intuito de promover a recuperação e a preservação da qualidade e quantidade dos recursos das bacias hidrográficas brasileiras, além da atuação na recuperação e preservação de nascentes, mananciais e cursos d'água em áreas urbanas (PORTAL EDUCAÇÃO, 2020).

De acordo com PORTAL EDUCAÇÃO (2020) um bom aproveitamento e gerencia-

mento dos recursos hídricos serve como um modo de combater aos efeitos adversos da poluição, da seca e do assoreamento. Adicionalmente, a gestão de recursos hídricos manejada como uma solução de otimização dos mesmos é capaz de gerar mais economia para indústrias.

O conceito de gestão integrada dos recursos hídricos implica na necessidade de uma boa gestão da água em função dos seus diferentes tipos de uso, das diferentes áreas de conhecimento envolvidas e dos diferentes tipos de instituições que utilizam e regularizam seu uso (PORTAL EDUCAÇÃO, 2020).

Ainda, é necessário que se olhe com atenção para as situações de risco ambiental e social, como enchentes e secas, sendo preciso ações preventivas das autoridades. Considerando a limitação desse recurso e a necessidade de compartilhamento a gestão integrada deve buscar uma tomada de decisão conjunta, onde os agentes envolvidos no problema interagem entre si para chegar em soluções mais justas e satisfatórias (ADAMATTI, 2007).

2.2 Jogos de Papéis e Jogos Sérios

Um RPG é um jogo que se situa na intersecção entre jogos e um teatro de improviso de modo que os jogadores interpretam uma personagem no contexto da história do jogo. Para cada papel é designado um conjunto de regras e ações que podem ou não ser feitos conforme a decisão do jogador que está interpretando-o. É esperado que os jogadores respeitem o escopo do limite de ações relativos ao seu papel no jogo.

Não há uma forma correta de se jogar RPG pois há diversas formas de se jogá-lo. O RPG mais tradicional é o conhecido RPG “de mesa”, onde os jogadores são interpretados através de diálogos. Em uma partida de RPG não há vencedores ou perdedores, já que o que se destaca não é seu aspecto competitivo, mas sim de cooperação. Os objetivos individuais ou coletivos devem ser buscados ao longo do jogo sempre respeitando as regras do jogo (ADAMATTI, 2007).

O RPG se destaca como uma técnica que vem sendo aplicada amplamente em treinamentos, pois permite-se colocar os jogadores em situações de tomada de decisão similares às reais, garantindo que não haja consequências efetivas. Grandes empresas têm feito o uso de RPG em cursos de treinamento devido ao fator lúdico envolvido nos jogos, o que faz com que o treinamento e/ou aprendizagem de determinado assunto seja facilitado (PERROTTON et al., 2017).

Outro aspecto que pode ser destacado do RPG é que pode servir como revelador de alguns aspectos das relações sociais, permitindo a observação direta das interações entre os jogadores (BARRETEAU; PAGE; D'AQUINO, 2003).

O intuito de um Jogo Sério (em inglês, *Serious Game*) na área educacional é pro-

porcionar um ambiente que crie a oportunidade de aprender, permitindo ao jogador ter experiências significativas, capaz de transmitir conhecimentos do jogo para o jogador (STAPLETON, 2004).

Os jogos sérios podem ser usados em um grande conjunto de áreas de aplicação, por exemplo, militar, governamental, educacional, corporativa e saúde (SUSI; JOHANNESSON; BACKLUND, 2007). Ainda existem discussões sobre o que o conceito em si significa. Entretanto, de acordo com Susi; Johannesson; Backlund (2007) a maioria dos autores concorda no significado geral de que jogos sérios são jogos usados para fins diferentes de mero entretenimento. O jogo sério possibilita aos alunos vivenciarem situações impossíveis no mundo real por razões de segurança, custo, tempo, e ainda permite que os jogadores desenvolvam diversas habilidades diferentes para jogar o jogo.

GORIM é o nome do jogo em desenvolvimento pelo grupo de pesquisa, onde esta dissertação se situa. Este é um jogo sério, do tipo RPG, que utiliza da técnica onde os jogadores “interpretam” uma personagem, criada dentro de um determinado cenário (ambiente). O ambiente do GORIM é um cenário de gerenciamento de recursos hídricos, focado na Lagoa Mirim e no Canal São Gonçalo, localizados nas cidades de Rio Grande e Pelotas no Rio Grande do Sul, ao sul do Brasil.

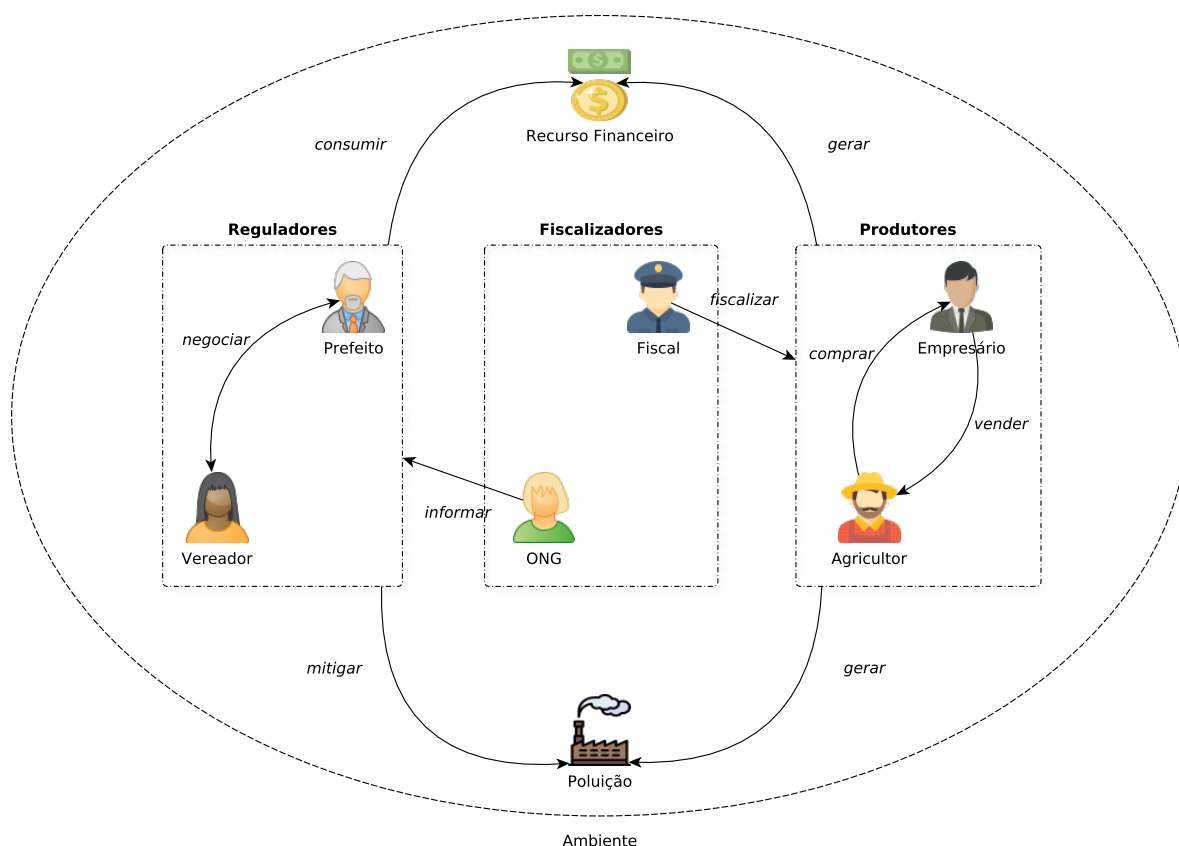


Figura 1 – Diagrama de agentes no ambiente do jogo. Fonte: (LEITZKE et al., 2019).

O intuito deste jogo é que os jogadores se comportem como agiriam em sua vida

real ao tomar decisões e ações no contexto do comitê de bacia. Os jogadores possuem papel deliberativo no ambiente de jogo, pois são capazes de tomar decisões em relação aos problemas sócio-ambientais apresentados.

O cenário do jogo distribui os jogadores como moradores das cidades de Atlantis e Cidadela. A modelagem de agentes está representada na Figura 1, mostrando os papéis que podem ser assumidos e como os agentes interagem no ambiente do jogo. Os agentes são classificados de acordo com o papel que assumem, os quais estão divididos em três grupos, são eles: reguladores, fiscalizadores e produtores.

Agentes reguladores incluem agentes como prefeitos e vereadores, cujo papel é da administração dos recursos financeiros de suas respectivas cidades. Esta verba pública é proveniente da arrecadação de impostos e taxas aplicadas aos demais agentes ao longo das rodadas. Prefeitos e vereadores podem chegar a acordos para aplicar ações que visam o controle e redução dos níveis de poluição, através de incentivos fiscais e tratamentos ao meio ambiente, sempre buscando não prejudicar os mecanismos de produção do ambiente do jogo. Ações permitidas aos prefeitos: aumentar ou diminuir taxas de impostos; uso de tratamento de água, esgoto ou lixo para diminuir a poluição; verificar propostas do vereador local. Ações permitidas aos vereadores: propor aumento ou diminuição de taxas de impostos.

O grupo dos agentes fiscalizadores é responsável por fiscalizar e/ou reportar irregularidades ligadas à exploração do meio ambiente, arrecadando assim as multas aos agentes que apresentam irregularidades. Estes valores vão diretamente para o caixa de sua respectiva prefeitura, como definido anteriormente. Em suma, o principal papel do agente fiscal é o de fiscalizar os agentes produtores, penalizando-os se violarem alguma regra imposta pelos agentes reguladores. Ações permitidas ao fiscal: inspecionar agricultores e empresários locais; multar agentes; dar ou retirar “selo verde”¹ para agricultores. A ONG representada no diagrama na Figura 1 tem sua modelagem feita através de um Personagem Não Jogável (em inglês, *Non-Player Character* – NPC)².

Por fim, os agentes produtores são os agentes que diretamente exploram o meio ambiente, esta exploração visa a obtenção de lucro. O grupo inclui empresários e agricultores. Os agentes empresários interagem com os agentes agricultores através da venda e/ou disponibilização de aluguel de insumos e equipamentos (que visam maximizar a produção).

Tanto os agricultores quanto os empresários são distribuídos entre Atlantis e Cidadela. Empresários de Semente e Máquina são moradores de Atlantis e Empresários de Fertilizantes e Agrotóxicos são moradores de Cidadela. Ações permitidas aos em-

¹ Refere-se a um selo que pode ser solicitado pelo agricultor caso sua plantação esteja livre de agrotóxicos. Além de ser um selo amigo do meio ambiente, o selo garante desconto em impostos.

² Personagens que fazem parte do cenário do jogo, realizando interações e não podem ser controlados por um jogador.

presários: vender/alugar³ produtos; negociação de valores. Ações dos agricultores: comprar insumos dos empresários; negociar com empresários; solicitar selo verde ao fiscal.

O jogo ocorre por rodadas, necessitando ter um número determinado de rodadas para sua realização. Cada rodada no jogo representa um ano de produção para os agentes. Cada rodada divide-se em duas etapas: a etapa dos agentes produtores (empresários e agricultores) e a etapa dos agentes reguladores e moderadores (prefeitos, vereadores e fiscais). A cada etapa os níveis de poluição são atualizado para os agentes, permitindo que os mesmos possam tomar decisões de acordo com a situação ambiental do momento. A Figura 2 demonstra os impactos do aumento da poluição no ambiente.

Níveis de Poluição



Figura 2 – Níveis de poluição do ambiente. Fonte: Manual do RPG GORIM⁴.

2.3 Aprendizado de Máquina

Métodos de Aprendizado de Máquina (AM) são usados nas áreas mais amplas, como serviços financeiros, aplicativos governamentais, marketing, transporte, bioinformática, medicina, astronomia, matemática estatística, teoria da informação, teoria dos jogos, otimização e outros. O AM é uma vertente específica da Inteligência Artificial (IA) que é capaz de treinar máquinas para aprender com os dados, suas aplicações devem ser capazes de aprender e se adaptar, tornando-se um meio comum

³O aluguel de produtos ocorre obrigatoriamente nas negociações com o empresário de máquina, pois as máquinas são apenas alugadas para o ano de produção devendo ser novamente locadas no próximo ciclo. Os demais empresários apenas vendem os insumos.

⁴Disponível em <https://t.ly/I3Qm>

para qualquer tarefa que necessite de reconhecimento de padrões significativos em grandes bases de dados (SHALEV-SHWARTZ; BEN-DAVID, 2014).

Abordando usos mais específicos no cotidiano é possível citar aplicações de algoritmos de buscas em bases de dados que aprendam constantemente qual a melhor resposta para o usuário ou, por exemplo, aprendizado para detecção de fraudes, protegendo as transações financeiras com cartões de crédito.

Nota-se que diferentemente da IA tradicional, cujo objetivo é a imitação automatizada de comportamento de inteligência, o foco do AM é usar os pontos fortes e habilidades dos computadores para complementar a inteligência humana, tarefas além das capacidades humanas. O objetivo é programar *softwares* capazes de aprender dada uma certa entrada. Executando efetivamente tarefas específicas sem uso de instruções explícitas, com uso de padrões e inferências (HARRINGTON, 2012).

Segundo Shalev-shwartz; Ben-david (2014) aprendizado é processo de incorporar experiências para que elas se tornem especialidades ou conhecimento. É esperado que as entradas no algoritmo sirvam como sua experiência para que seja possível gerar inferências, ou seja, conhecimento na sua saída. Contudo, Alpaydin (2020) define o AM como a tarefa de programar computadores para melhorar os critérios de desempenho usando a experiência.

O conhecimento do algoritmo não deve se limitar ao seu treinamento inicial, assim como os humanos, um bom algoritmo pode por toda a sua execução estar em constante aprendizado conforme novas entradas são inseridas e aprendendo sobre seus erros (EL NAQA; MURPHY, 2015). Existe um limiar delicado sobre o quão grande deve ser o conhecimento prévio do algoritmo, conhecimento que serve de base para o processo de aprendizado. Evidentemente, existe uma relação entre a proporcionalidade de quanto mais se aprende mais fácil se torna aprender com outros exemplos. Todavia, se o conhecimento inicial, ou seja as premissas que o algoritmo assume como verdadeiras forem extremamente fortes pode gerar por consequência uma inflexibilidade na aprendizagem, já que esta é *a priori* vinculada com as premissas iniciais que foram inseridas.

Existem quatro subcategorias de aprendizado de máquina: aprendizado supervisionado, aprendizado não supervisionado, aprendizado semi-supervisionado e aprendizado por reforço (HURWITZ; KIRSCH, 2018).

No aprendizado supervisionado é preciso uma entrada para treinamento dos dados, ou seja, é necessário um conjunto de dados que possa ser usado como base de aprendizado para a eventual classificação de novas entradas gerando um sistema que aprende sob um conjunto de dados predefinido. Assim, o sistema pode informar para aquele conjunto qual a saída correta. Portanto, gera-se um padrão que permite a leitura de novas entradas com suas saídas determinadas corretamente. A proposta é construir um modelo de aprendizado de máquina a partir desses pares de

entrada/saída, formando um conjunto de treinamento e usá-lo para fazer previsões precisas para novos dados nunca antes vistos como se refere (MULLER; GUIDO, 2017). Um exemplo de aprendizado supervisionado é a previsão do tempo, utilizando-se da análise de regressão são considerados os padrões meteorológicos registrados e as condições atuais para que o tempo possa ser previsto (HURWITZ; KIRSCH, 2018).

Na aprendizagem não supervisionada, o algoritmo busca encontrar padrões sem qualquer conhecimento histórico sobre os dados. De acordo com Muller; Guido (2017) não há saída conhecida, nenhuma resposta correta informada ao sistema ou nenhum professor para instruir o algoritmo de aprendizagem, são apenas mostrados os dados de entrada e solicitados a extrair conhecimento desses dados. O objetivo é explorar os dados informados e descobrir alguma estrutura nestes dados. O aprendizado ocorre sem conhecimento prévio, assim o algoritmo não tenta obrigatoriamente associar cada entrada à um rótulo (como ocorre no aprendizado supervisionado). O algoritmo aprende sobre os dados de entrada inseridos. Uma aplicação da aprendizagem não supervisionada seria na área da saúde, pois é uma área que tem uma grande coleta de dados, possibilitando a abordagem de rotular as fontes de dados a uma doença específica (HURWITZ; KIRSCH, 2018).

O aprendizado semi-supervisionado é a união dos dois tipos de aprendizados anteriores, onde uma parte dos dados possui uma classificação e outra não. Assim a parte já classificada pode ser usada para auxiliar na classificação da parte não classificada. Esta aprendizagem manipula dados de rótulo e nenhum dado de rótulo para treinamento, é um paradigma de aprendizagem em que o objetivo é aprender a partir de ambos os dados (rotulados e não rotulados). Normalmente, há uma pequena quantidade de dados rotulados e a maior quantidade de dados não rotulados, uma vez que dados não rotulados são frequentemente comuns para obter acesso. O ponto principal da aprendizagem semi-supervisionada é explorar a combinação de dados rotulados e não rotulados, entendendo como ela pode inferir o processo de aprendizagem e desenvolver algoritmos que aproveitem tal combinação (ZHU; GOLDBERG, 2009). Este tipo de processo de aprendizagem é útil ao obter todos os dados rotulados (e aplicar um treinamento rotulado), porém é um processo muito caro.

Por último, o aprendizado por reforço é definido por Kaelbling; Littman; Moore (1996) como “o problema enfrentado por um agente que aprende o comportamento por meio de interações de tentativa e erro com um ambiente dinâmico”. O algoritmo busca por ações que gerem melhores recompensas. Muito utilizado em robótica e em jogos, já que é amplamente utilizado em treinamento de robôs. Um exemplo de situação seria um treinamento de navegação de robô em um lance de escadas. O robô constantemente aprende com suas ações, caso ele caia, os dados são recalculados para que o robô navegue com um procedimento diferente até que seu treinamento esteja completo. O robô é treinado por tentativa e erro para aprender como subir

escadas (HURWITZ; KIRSCH, 2018).

Quatro modelos de aprendizagem supervisionada são utilizados neste trabalho. Os modelos produzidos durante este estudo tratam-se de modelos de regressão. Os algoritmos selecionados foram: Regressão Linear⁵, *Decision Tree Regression*⁶, *Random Forest Regression*⁷ e *Support Vector Regression*⁸.

A Regressão Linear é um dos modelos mais usados para prever resultados (MONTGOMERY; PECK; VINING, 2012). A análise dos resultados obtidos com modelos de regressão compreende o estudo do relacionamento entre variáveis dependentes e outras variáveis chamadas independentes. A equação linear representa a relação entre a variável dependente do modelo e um conjunto de variáveis independentes. A Regressão Linear pode ser apresentada em duas maneiras: simples e múltipla.

A Regressão Linear Simples é composta por uma única variável independente no processo de previsão, conforme definido na Equação 1 (MONTGOMERY; PECK; VINING, 2012).

$$y = \beta_0 + \beta_1 x + \epsilon \quad (1)$$

onde y é uma variável dependente, β_0 é uma constante, x é uma variável independente com um coeficiente β_1 .

No caso da Regressão Linear Múltipla, há mais de uma variável independente envolvida no processo. A resposta y pode ser relacionada a k variáveis regressoras. O modelo é representado na Equação 2 (MONTGOMERY; PECK; VINING, 2012).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (2)$$

O conceito por trás do modelo de *Decision Tree Regression* envolve a compreensão sobre Árvores de Decisão. Árvores de Decisão apresentam resultados satisfatórios quando aplicadas a problemas de regressão. O processo de decisão ocorre por etapas hierárquicas ou como um esquema de árvore. A estrutura da árvore segue conforme a composição: um nó raiz, que concentra todos os dados; um conjunto de nós internos (representando divisões); e, um conjunto de nós terminais, ou seja, as folhas da árvore. Na abordagem de cima para baixo, cada nó da estrutura toma uma decisão binária que separa uma classe ou um conjunto de classes das demais. Nesta abordagem, o processamento ocorre no sentido para baixo da árvore, sempre buscando alcançar o nó folha (XU et al., 2005).

⁵https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

⁶https://scikit-learn.org/stable/auto_examples/tree/plot_tree_regression.html

⁷<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

⁸<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>

Essencialmente uma árvore de decisão busca dividir uma decisão complexa em múltiplas decisões simples, atingindo uma solução, no nó folha. Nesta abordagem, os dados representam as variáveis preditoras e a classe que se deseja mapear representa a variável de saída (BREIMAN et al., 1984). A seleção e classificação de recursos são realizadas simultaneamente, visto que recursos com mais informações disponíveis são selecionados automaticamente para a classificação e os demais rejeitados, o que assegura um aumento da eficiência computacional. Ademais a estrutura em três da árvore de decisão torna possível a compreensão e interpretação das informações em cada nível do processo.

O modelo *Decision Tree Regression* é de fato uma variação do classificador de Árvore de Decisão (do inglês, *Decision Tree Classifier*), com objetivo de aproximar funções de valor real, como as proporções de classe. A construção de uma árvore de regressão também é baseada no particionamento recursivo binário, que é um processo iterativo que divide os dados em partições. Em um primeiro momento todos os exemplos de treinamento são considerados para a determinação da estrutura da árvore. Assim, o algoritmo separa os dados através das divisões binárias possíveis, buscando a divisão responsável por dividir os dados em duas partes, de modo que a soma dos desvios quadrados da média nas partes separadas é minimizada. O processo de divisão é novamente aplicado a cada um dos novos ramos e ocorre até que cada nó atinja um tamanho mínimo, tornando-se um nó terminal (BREIMAN et al., 1984). De modo geral, o algoritmo é semelhante à estrutura de uma árvore invertida com os valores preditivos localizados nos nós terminais (NISBET; ELDER; MINER, 2009).

O modelo *Random Forest Regression* utiliza-se de Florestas Aleatórias, que representam uma combinação de Árvores de Decisão. Em uma Floresta Aleatória, cada nó é dividido usando o melhor divisor possível dentre o subconjunto de todos os divisores. (BREIMAN, 2001) destaca que apesar de ser uma estratégia contra-intuitiva, se demonstra robusta contra *overfitting*⁹. O diferencial das Florestas Aleatórias é que este subconjunto de divisores é gerado aleatoriamente e então explorado pelo algoritmo, o que não gera uma busca por um divisor ideal de um determinado nó ao avaliar-se todas as divisões permitidas em todas as covariáveis (SEGAL, 2003). O algoritmo *Random Forest Regression* utiliza-se do processo de votação, levando em consideração diferentes árvores de decisão e lhes atribuindo o processo de decisão. A regressão nas florestas aleatórias pode ter sua resposta final determinada como a média dos resultados de todas as previsões das diferentes árvores.

As Máquinas de Vetores de Suporte (do inglês, *Support Vector Machine – SVM*)

⁹ *Overfitting* ocorre geralmente em modelos de aprendizado supervisionados pois os os modelos são treinados em um subconjunto de dados rotulados. Nesta situação o modelo se ajusta extremamente bem ao conjunto de dados, porém se demonstra incapaz de generalizar para novos dados.

são modelos de aprendizado supervisionado que se aplica em modelos de classificação e regressão. *Support Vector Regression* (SVR) é o termo correto a ser aplicado tratando-se de modelos de regressão. SVR pode ser aplicado em problemas de regressão linear ou não linear dependendo da função de kernel aplicada.

De acordo com Kavitha; Varuna; Ramya (2016) o *Support Vector Machine* utiliza-se de um kernel linear nas funções de regressão, de modo semelhante ao SVM. O SVR define sua margem de tolerância (ε) por aproximação, utilizando-se de conceitos do SVM para a solução de problemas de regressão, buscando pela margem máxima que separa o hiperplano (AWAD; KHANNA, 2015). A largura entre o hiperplano (conceito generalizado de reta ou plano para várias dimensões) e as margens é controlada pela tolerância ε .

São utilizadas determinadas métricas para avaliar este conjunto de modelos: o coeficiente de determinação (do inglês, *coefficient of determination* – R^2), o erro médio absoluto (do inglês, *Mean Absolute Error* – MAE) e o erro médio quadrado (do inglês, *Mean Squared Error* – MSE). Cada métrica de avaliação utilizada fornece informações sobre o modelo.

O coeficiente de determinação (R^2) é uma medida utilizada para analisar as relações diretas de diferença entre uma variável determinada por uma segunda variável. O R^2 tem seu intervalo entre 0 e 1, também pode ser expresso em termos percentuais. Seu percentual determina a quantidade da variância dos dados que é demonstrada pelo modelo linear. Logo, quanto maior valor o R^2 , melhor ele se ajusta à amostra. O R^2 , assemelha-se ao coeficiente de correlação, R. A proposta do coeficiente de correlação é determinar o quão forte é uma relação linear entre duas variáveis. R^2 é o quadrado do coeficiente de correlação (GLEN, 2015).

MAE é uma medida para calcular a média do erro absoluto que ocorre entre o valor previsto e o valor real. É uma média de todos os erros absolutos, capaz de medir o quão longe o modelo está de prever os valores corretos (GLEN, 2016). MSE funciona de modo semelhante, mas os erros são elevados ao quadrado antes que a média seja calculada, demonstrando o quão próxima a linha de regressão (gerada pelos modelos) está de um determinado conjunto de pontos (classe prevista). Portanto, quanto menor for o MSE, significa que mais perto a regressão está de representar o modelo com precisão (GLEN, 2013).

2.4 Análise de Sentimentos

A Análise de Sentimentos (AS) refere-se ao uso de Processamento de Língua Natural (PLN), Linguística Computacional e Análise de Texto para identificar, extrair, quantificar e estudar estados afetivos e informações subjetivas. O principal objetivo da AS é definir técnicas automáticas capazes de extrair informações subjetivas de textos,

como opiniões e sentimentos, a fim de criar conhecimento estruturado que possa ser utilizado por um sistema de apoio ou tomador de decisão. A AS é uma das áreas de pesquisa mais destacadas em PLN desde o início dos anos 2000, quando se tornou uma área de pesquisa muito ativa (LIU, 2010).

Tratando-se dos modos possíveis de se realizar esta análise, os estudos de Mejova (2009) e de Feldman (2013) classificam a AS em três níveis de granularidade (do maior ao menor): a análise do documento de modo integral, a análise ao nível de sentença e a análise baseada em aspecto.

De acordo com Feldman (2013), na AS ao nível do documento assume-se que o documento expressa uma opinião de maneira geral, um ponto de vista deixado pelo autor. A AS ao nível de sentença pressupõe que um documento pode conter diversas opiniões. Feldman (2013) elucida que ao buscar uma informação mais detalhada sobre as múltiplas opiniões contidas em um documento passa-se a realizar uma análise sob um menor nível, atribuindo-se uma opinião particular a cada sentença. Na AS ao nível de aspectos, dado um documento procura-se reconhecer no mesmo os aspectos e todas as expressões de sentimentos referentes a estes aspectos.

O trabalho intitulado *A Survey on Sentiment and Emotion Analysis for Computational Literary Studies* desenvolvido por (KIM; KLINGER, 2018b) tem como objetivo demonstrar uma ampla visão dos métodos utilizados na análise de emoção e sentimento aplicados a um texto. O estudo acerca das emoções inicia com a pesquisa de Darwin; Prodger (1998) sendo até hoje vasto e diverso. A pesquisa se concentra em três pontos de vistas comumente usados para análise computacional: A teoria das emoções básicas de Ekman, a roda da emoção de Plutchik e o modelo circumplexo de Russell.

A teoria de Ekman (EKMAN, 1993) tem sua fundamentação na observação das expressões faciais no início do desenvolvimento ou interação social. A teoria de Plutchik (PLUTCHIK, 1991) diferencia-se principalmente da teoria de Ekman por além abordar sobre um conjunto de emoções, representa que todas as outras emoções são misturadas e portanto derivadas das várias combinações das emoções do conjunto básico. Ademais, Plutchik representa essas emoções em uma estrutura determinada “roda das emoções de Plutchik”.

A teoria de Russell é representada em seu modelo circumplexo¹⁰, que é cortado nos seus eixos horizontais e verticais, dividindo as emoções por dimensões, pois as emoções propostas por Russell são concentradas na periferia de um círculo. Analisando diversos estudos e comparando as abordagens entre eles (KIM; KLINGER, 2018b) destaca que a principal mudança é na metodologia onde os estudos iniciais

¹⁰O circumplexo é uma representação gráfica circular que utiliza de coordenadas polares capaz de representar as semelhanças entre múltiplas variáveis (GIGANDET, 2015). A interpretação do modelo permite compreender que a correlação entre as variáveis diminui à medida que a distância entre eles no círculo aumenta (ASSOCIATION, 2020)

de análise de sentimento baseavam-se na polaridade das palavras e em dicionários específicos. As últimas pesquisas utilizam fortemente *word embeddings*¹¹ e aprendizado profundo. Contudo a maioria dos estudos ainda utiliza técnicas de contagem de palavras ou recursos léxicos. (KIM; KLINGER, 2018b) estima que cerca de 70% dos trabalhos abordados no estudo dependem substancialmente do uso de vários recursos lexicais para detecção de emoções, assim trabalhos recentes ainda baseiam sua análise em dicionários ou contagem de palavras.

Os estudos na área de AS registram duas principais estratégias para que esta análise ocorra. Neethu; Rajasree (2013) destaca a abordagem por AM e a abordagem por léxicos.

Em estudos como os de Pang; Lee; Vaithyanathan (2002) e de Neethu; Rajasree (2013), são utilizadas as técnicas de AM para a classificação de sentimentos. Segundo Pang; Lee; Vaithyanathan (2002), a dificuldade na aplicação de técnicas de AM deve-se ao fato de que em uma classificação tradicional, por exemplo, palavras-chaves são capazes de determinar as saídas esperadas, porém, tratando-se de sentimentos tem-se algo expresso de um modo mais subjetivo.

Algoritmos de AM utilizam-se de conjuntos de treino e teste, onde no conjunto de treino os dados de entrada são vetorizados e associados a uma saída específica. Assim é esperado que, através do conjunto de teste, seja possível classificar outros vetores de entrada nas saídas corretas (NEETHU; RAJASREE, 2013). Alguns recursos são comumente usados para que ocorra a classificação de sentimentos como Frequência de Termos, Presença de Termo, Negação, *N-Grams* e *Part-Of-Speech* (MEJOVA, 2009).

Nas análises realizadas por Araújo et al. (2013) foram utilizados diferentes léxicos de sentimentos disponíveis. Carvalho; Silva (2015) explica que as aplicações de análise de sentimentos baseiam-se em léxicos de sentimentos para gerar seus resultados, sendo estes definidos como “léxicos cujas entradas podem ser utilizadas para veicular um determinado sentimento ou emoção”.

De acordo com Feldman (2013), existem três modos de se adquirir um léxico de sentimento: i) produção manual do léxico; ii) produção de um léxico baseado em dicionário através de um conjunto de palavras expandidas por uma base de dados; e, iii) aquisição de um léxico através da uma expansão de um conjunto inicial de palavras em um grande corpus. Neethu; Rajasree (2013) destaca em seu trabalho sobre a dificuldade da realização da AS através de léxicos devido a exigência de uma grande base de dados.

¹¹ *Word Embeddings* são um conjunto de técnicas que permitem que palavras individualmente sejam representadas como vetores de valor real em um espaço vetorial predefinido. Mapeia-se cada palavra para um vetor e através de um modo semelhante a uma rede neural, o restante dos valores do vetor são aprendidos (BROWNLEE, 2019).

2.5 Teorias de Emoções

Neste trabalho a análise de sentimento realizada foi desenvolvida utilizando-se das teorias das emoções abordadas do estudo de Kim; Klinger (2018b). A classificação de emoções é uma tarefa desafiadora capaz de ser computacionalizada. Apesar de existirem diferentes teorias acerca das emoções, de modo geral os teóricos concordam que as emoções englobam um conjunto de características expressivas, comportamentais, fisiológicas e fenomenológicas. O estudo se concentra em analisar algumas teorias das emoções já amplamente conhecidas na análise computacional, para este trabalho o estudo se concentra na teoria das emoções básicas de Ekman (EKMAN, 1993) e na roda das emoções de Plutchik (PLUTCHIK, 1991).

A teoria proposta por Paul Ekman considera que as manifestações faciais das emoções fundamentais não são aprendidas, mas são inerentes aos seres humanos. Contudo, ainda há limitações culturais sobre como e em quais situações cada emoção se aplica. As emoções básicas apresentadas por Ekman são: alegria, medo, nojo, raiva, surpresa e tristeza. A teoria de Ekman baseia-se principalmente na observação do comportamento facial, no início do desenvolvimento ou interação social. Ainda, considera-se que as emoções possuem intensidades diferentes e são independentes, ou seja, as emoções não se misturam (EKMAN, 1993).

O modelo tridimensional desenvolvido por Robert Plutchik descreve as relações entre as emoções (DONALDSON, 2017), classificando-as em oito categorias principais, sendo metade destas emoções positivas e outra metade negativa (MOHSIN; BELTIUKOV, 2019). Assume-se que qualquer emoção possa ser expressa por alguma dessas oito emoções, sendo estas: alegria, antecipação, confiança, medo, nojo, raiva, surpresa e tristeza. Plutchik ainda ressalta que estas emoções são culturalmente independentes, possibilitando a aplicação desta teoria independente de cultura ou idioma (TROMP; PECHENIZKIY, 2014). De acordo com Donaldson (2017), a roda das emoções de Plutchik auxilia não só no reconhecimento das emoções, mas na compreensão de como diferentes emoções estão relacionadas entre si e como podem se modificar com o tempo.

2.6 Trabalhos Relacionados

Esta Seção descreve alguns trabalhos relacionados encontrados na literatura apresentados em (KESHTKAR et al., 2012), (KANSAON; BRANDÃO; PINTO, 2018) e (GAIND; SYAL; PADGALWAR, 2019), respectivamente. A quantidade reduzida de estudos abordados deveu-se ao fato de que buscou-se focar em estudos que apresentassem abordagens diferentes acerca do tema de análise de sentimentos. É importante salientar que todos estes trabalhos são brevemente apresentados e, então, faz-se uma análise comparativa dos mesmos.

A ideia de se detectar traços de personalidade e comportamento através de conversas não é uma tarefa fácil. No estudo de Keshtkar et al. (2012), intitulado *Detecting Players Personality Behavior with any Effort of Concealment*, é abordado o uso de técnicas de PLN e de técnicas de AM considerando o diagrama de Leary's Rose para a predição de personalidades de jogadores no jogo *Land Science*.

Este jogo foi projetado para simular um planejamento regional de experiência prática para os alunos, onde os jogadores atuam como estagiários em uma empresa de planejamento urbano e discutem sobre as situações da empresa. Estas discussões de ideias e estratégias produzem os textos em língua natural. O uso do diagrama de Leary's Rose permite que ocorra a classificação da personalidade dos jogadores em seis possíveis classes, são elas: competitivo, líder, dependente retirado, oposto e juntos. Os textos utilizados no experimentos foram obtidos através do *chat* de conversa do jogo e para as 1.000 entradas houve uma classificação manual em uma das seis classes.

Para processamento desses textos foram usados quatro diferentes léxicos de sentimento: SentiWordNet (ESULI; SEBASTIANI, 2006), Subjectivity lexicon (WIEBE; WILSON, 2002), General Inquirer (GILMAN, 1968) e Taboada (TABOADA; GRIEVE, 2004). O *software* Linguistic Inquiry and Word Count (LIWC) (PENNEBAKER; BOOTH; FRANCIS, 2007) é comumente usado para a detecção de traços de personalidade. Porém como o LIWC não possui um classificador de texto, foram criadas *features* com base na saída do LIWC. Por fim, para a classificação de sentimento foram usados os seguintes classificadores: *Naive Bayes* (NB), Máquina de Vetores Suporte (SVM) e a Árvore de Decisão (J48), sendo todos executados através da ferramenta *Weka* (HOLMES; DONKIN; WITTEN, 1994)¹².

O teste de cada classificador ocorre através do uso de *unigrams*¹³, *bigrams*, *unigrams+bigrams*, *LIWC+bigrams*, *lexicons+bigrams* e *unigrams+LIWC*, sendo a mais alta acurácia obtida, de 83,71%, do modelo SVM com uso de *lexicons+bigrams*.

O estudo de Kansaon; Brandão; Pinto (2018), intitulado Análise de Sentimentos em *Tweets* em Português Brasileiro, tem como objetivo encontrar quais algoritmos de classificação podem melhor descobrir sentimentos em uma análise de textos curtos providos de uma rede social.

Os dados foram coletados do *Twitter* com restrição de estar em português brasileiro e possuir as *hashtags* definidas (#Triste, #Chateado, #Feliz, #Amor, #Raiva, #Inveja, #Ironia), que representam respectivamente os sentimentos escolhidos para serem analisados. A coleta inicial contou com 12.814 *tweets* porém após uma sequência de etapas de pré-processamento, que incluíam: remoção de *tweets* repetidos, re-

¹² *Weka* é um *software* multiplataforma que contém uma coleção de algoritmos ou abordagens de aprendizado de máquina para tarefas de mineração de dados.

¹³ *N-grams* representam um conjunto de palavras consecutivas. Um *n-gram* de tamanho 1 é conhecido como um *unigram*; o de tamanho 2 por *bigram* e o de tamanho 3 por *trigram* (FUEYO, 2018).

moção das *hashtags* da busca e outras informações retornadas pela busca que não seriam utilizadas, o número final de *tweets* analisados foi de 9.631.

Ademais, para gerar um bom balanceamento dos sentimentos utilizados, definiu-se que para um sentimento ser analisado deveria conter no mínimo mil *tweets* na base de dados e, por isso, sentimentos como Ironia, Inveja e Raiva foram removidos do estudo. Na prática, com o número de 9.631 *tweets* e quatro sentimentos, estabeleceu-se uma média de cerca de dois mil *tweets* para cada sentimentos.

Os algoritmos de classificação foram selecionados baseados em modelos *Naive Bayes* ou em modelos que apresentaram bons resultados em estudos analisados em Kansaon; Brandão; Pinto (2018). Sendo estes *Naive Bayes*, *Naive Bayes Multinomial*, *Naive Bayes Multinomial Updateable*, *Sparge Generative Model*, *Discriminative Multinomial Naive Bayes Text*, *Complement Naive Bayes*, *Bayesian Logistic Regression*, *Instance-Bases Learning Witj Parameter K*, *Forest and Random Committee*. A execução e treinamento ocorreu através da ferramenta *Weka* (HOLMES; DONKIN; WITTEN, 1994).

O método utilizado para treinamento foi a validação cruzada, processo que visa avaliar por completo a precisão do modelo de forma prática através da realização de diversas separações de conjuntos de treino e teste em vários cenários de amostragem, permitindo ao estudo a obtenção de resultados melhores. Os algoritmos baseados no modelo *Naive Bayes* obtiveram de modo geral os melhores resultados, em especial o *Complement Naive Bayes* que atingiu por volta de 85% de acerto na classificação.

O estudo de Gaiind; Syal; Padgalwar (2019), intitulado *Emotion Detection and Analysis on Social Media*, detecta emoções em *tweets* no idioma Inglês. A coleta dos dados ocorre via *Tweepy*¹⁴, uma biblioteca do Python para a coleta de *tweets* e o primeiro processamento é remover duplicatas, *links*, *hashtags* e palavras que não estejam escritas em inglês.

O *Tweets Set* gerado por esta coleta resultou em uma base com cerca de 10.000 entradas. Outras bases de dados são usadas no estudo, como *Emotion-Words Set* (EWS), *Degree-Words Set* (DWS) e *Location-Areas Set* (LAS). O EWS é um conjunto de palavras em torno de 1.500 palavras desenvolvido pelo estudo por busca em profundidade (DFS – do inglês, *Depth-First Search*).

A DFS busca os sinônimos das 6 categorias básicas de emoção em até dois níveis, sendo estas: felicidade, tristeza, raiva, surpresa, medo e nojo, em um dicionário de sinônimos. Por fim, cada palavra foi classificada manualmente em uma categoria de intensidade (fraco, médio ou forte). O DWS inclui palavras que são consideradas intensificadores, as quais podem atribuir mais força ou diminuir a intensidade da emoção descrita na sentença. Assim, cada palavra armazenada tem um grau de intensidade associado a ela. O LAS armazena as áreas de cerca de vinte grandes cidades da

¹⁴<https://www.tweepy.org/>

Índia.

O trabalho desenvolveu dois modelos completamente diferentes, os quais estão correlacionados. O primeiro modelo faz uso de PLN e o segundo modelo faz uso de AM. O modelo de PLN é desenvolvido da seguinte forma: i) usa o conjunto EWS sob os *tweets* para encontrar palavras em comum e marcá-las¹⁵; ii) identifica a pessoa, se a publicação é em primeira pessoa ou se refere a uma terceira pessoa; iii) considera o efeito de intensidade ou negação sob uma emoção, por exemplo, caso uma palavra tenha sido marcada por pertencer ao EWS mas acompanhar uma negação previamente muda o sentimento; iv) detecta *emoticons* que já estão tabelados no EWS; e, v) por fim, é calculada uma pontuação relativa a cada emoção analisada.

O modelo de AM faz uso da ferramenta *Weka* (HOLMES; DONKIN; WITTEN, 1994) e dos classificadores: Otimização Mínima Sequencial (SMO) e Árvore de Decisão (J48) aplicados no *Tweets Set*. Estes classificadores obtiveram ótimos resultados, com acurácia de 91,7% e 85,4%, respectivamente.

Tabela 1 – Comparação entre os trabalhos relacionados.

Trabalho	Algoritmos de Classificação	Ferramenta	Datasets	Léxicos
(KESHTKAR et al., 2012)	<i>Naive Bayes</i> (NB), Máquina de Vetores Suporte (SVM) e Árvore de Decisão (J48)	Weka	<i>Chat</i> de conversa do jogo <i>Land Science</i>	SentiWordNet, Subjectivity lexicon, General Inquirer, Taboada e LIWC
(KANSAON; BRANDÃO; PINTO, 2018)	Algoritmos Baseados no Modelo <i>Naive Bayes</i>	Weka	Twitter	Não utiliza
(GAIND; SYAL; PADGALWAR, 2019)	Otimização Mínima Sequencial (SMO) e Árvore de Decisão (J48)	Weka	Twitter	<i>Emotion-Words Set</i> (EWS), <i>Degree-Words Set</i> (DWS) e <i>Location-Areas Set</i> (LAS)

Os estudos citados nesta Seção fazem uso de diferentes técnicas e abordagens de AS, porém a forma como isto ocorre é particular de cada estudo. A Tabela 1 auxilia a visualizar as diferenças e semelhanças acerca deles. Os algoritmos de classificação escolhidos são diferentes para cada estudo, em Kansaon; Brandão; Pinto (2018) são apenas escolhidos algoritmos baseados no modelo *Naive Bayes*. As Árvores de Decisão (J48) foram avaliadas nos estudos de Keshtkar et al. (2012) e de Gaind; Syal; Padgalwar (2019). É importante salientar, que todos os modelos são construídos através do uso da ferramenta *Weka* (HOLMES; DONKIN; WITTEN, 1994).

Os estudos de Kansaon; Brandão; Pinto (2018) e de Gaind; Syal; Padgalwar (2019) realizam AS com foco em redes sociais, mais especificamente o *Twitter*. Já o trabalho de Keshtkar et al. (2012) tem sua análise voltada para um contexto de jogos, utilizando-se do *chat* de conversa do jogo, assemelhando-se a essa dissertação, cujos dados são provenientes das conversas do GORIM.

Ainda que a maioria dos estudos foquem em um classificador de sentimentos através de modelos de AM, os estudos de Keshtkar et al. (2012) e de Gaind; Syal; Pad-

¹⁵A palavra é considerada como um acerto, porém a pontuação somente é calculada ao fim da realização de todas as etapas do modelo de PLN

galwar (2019) utilizam-se de léxicos de sentimento e/ou outras bases de dados auxiliares. O trabalho apresentado aqui também depende substancialmente do uso de léxicos de sentimento.

3 ABORDAGEM PROPOSTA

Neste Capítulo são descritos o processo de construção dos *datasets*, do léxico de sentimento e do dicionário de termos intensificadores, do classificador de sentimentos e dos modelos de preditores.

O processamento e pré-processamento foi desenvolvido na Linguagem *Python* 3.7¹ (LUTZ, 2001) através da plataforma *Jupyter Notebook*². Ainda, bibliotecas como *Pandas*, *Numpy*, *Scikit-Learn* e *spaCy* foram utilizadas no desenvolvimento deste trabalho.

3.1 Construção dos *Datasets*

Nesta Seção serão apresentadas as etapas realizadas para a construção do *dataset* oriundo do motor do RPG GORIM e dos diálogos entre os personagens fazendo as transações.

3.1.1 Dados do motor RPG GORIM

O *dataset* oriundo do motor do RPG GORIM foi coletado em cinco partidas ocorridas totalmente online ao longo do ano de 2020. As partidas ocorreram no seguinte padrão: em cada partida houve de quatro a cinco rodadas com 8 ou 10 jogadores. As ações de cada rodada foram registradas em planilhas do Google, permitindo a inserção destes dados no motor do RPG GORIM que processa as ações de cada jogador. A saída final produzida pelo motor do RPG GORIM resulta em um compilado de todas as ações de todos os jogadores ao longo da partida.

A saída final produzida pelo motor do RPG GORIM resulta em um *dataset* não padronizado, sendo preciso reorganizar os resultados como registro de ações entre jogadores.

¹<https://www.python.org/>

²<https://jupyter.org/>

A Tabela 2 mostra exemplo do dado de uma transação gerado pelo motor RPG GORIM. Já a Tabela 3 mostra exemplo da mesma transação com os dados reorganizados.

Tabela 2 – Exemplo de dado gerado pelo motor RPG GORIM.

venda/aluguel	EmpSem	AgAT2	1	Soja	normal
---------------	--------	-------	---	------	--------

Para obter o *dataset* oriundo do motor do RPG GORIM padronizado, foi necessário alinhar todas as informações ou ações que são permitidas aos agentes fazerem em uma rodada. Portanto, cada linha representa uma ação de um agente ou uma transição entre dois agentes. Ademais foi necessário recalculer alguns dos custos, impostos e saldos para cada transação, permitindo obter-se o valor da poluição por transação.

O conjunto de dados final se assemelha com os dados apresentados na Tabela 3. É importante salientar que esta Tabela toda corresponde a uma linha do *dataset* produzido.

Tabela 3 – Exemplo dos dados reorganizados.

ação	agente1	agente2	parcela	objeto
venda	EmpSem FiscalAT	AgAT2 VereadorAT	1.0	Soja
tipo preço	rodada	jogo	saldo agricultor CD1	saldo agricultor CD2
normal	1	5	600.0	600.0
saldo agricultor CD3	saldo agricultor AT1	saldo agricultor AT2	saldo agricultor AT3	saldo EmpAgro
600.00	600.0	600.0	1000.00	1000.00
saldo EmpFert	saldo EmpMaq	saldo EmpSem	saldo FiscalAT	saldo FiscalCD
1000.00	1000.00	1000.00	0.00	0.00
saldo PrefeitoAT	saldo PrefeitoCD	poluição mundial	selo	próxima taxa
1000.00	1000.00	20.33	NaN	NaN
taxa	tratamento			
NaN	NaN			

Em um segundo momento foi realizado o pré-processamento do *dataset* para evitar a geração de erros futuros ou entradas inválidas para a análise. Ajustou-se valores com erros ou nulos e também foram feitas alterações na nomenclatura utilizada no nome dos agentes para que houvesse uma padronização ao longo de todo o estudo. Estas etapas e as seguintes que sumarizam o processo de modificação do *dataset* oriundo do motor RPG GORIM para o novo *dataset* (dados reorganizados) podem ser compreendidas na Figura 3.

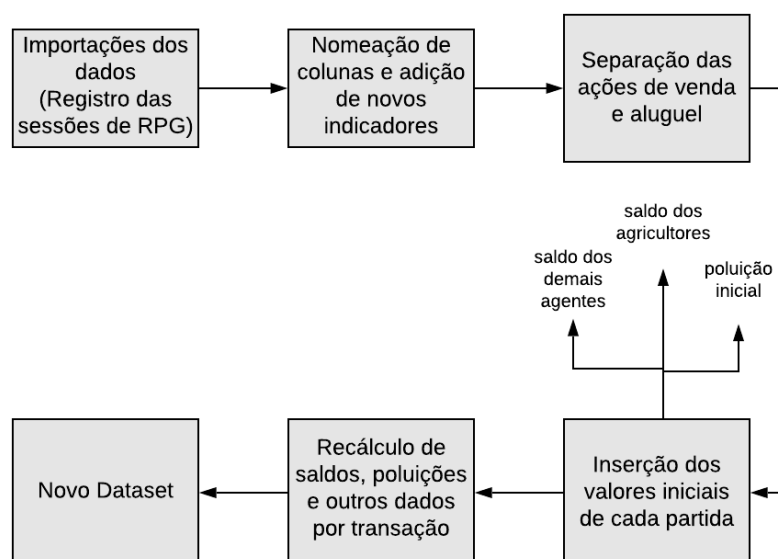


Figura 3 – Diagrama das etapas realizadas para a construção do novo *dataset*. Fonte: Própria.

3.1.2 Diálogos entre os personagens fazendo as transações

Um dos corpus construídos durante a elaboração deste trabalho é composto por diálogos entre os personagens durante as partidas do RPG GORIM. As conversas e as negociações dos jogos ocorreram através do *chat* do *Discord*³ (CARDOSO, 2018) utilizando-se exclusivamente da comunicação escrita para facilitar a coleta de dados. No servidor do *Discord* disponibilizado aos jogadores há uma sala de negociação para cada empresário. Os diálogos das negociações foram manualmente copiados e reorganizado em transações, onde por uma transação entende-se que ocorreu uma negociação entre um empresário e um agricultor.

Os diálogos inicialmente copiados isolados continham um registro semelhante ao demonstrado na Figura 4. Este trecho em específico registra o diálogo entre o empresário de máquinas e o agricultor de Atlantis 1 na primeira rodada de uma das partidas do GORIM. O Apêndice A contém um registro integral dos diálogos entre o empresário de máquinas e todos os agricultores com quem realizou negociações na primeira rodada de determinada partida.

Assim como nos dados provindos do motor do RPG GORIM foi necessário um pré-processamento para ajuste das nomenclaturas utilizadas pelos jogadores ao longo das partidas. O apelido dos usuários é alterado para o padrão Agente Etapa 1|Agente Etapa 2, no caso do jogador ocupar um cargo de agente fiscalizador ou regulador, caso o jogador esteja no papel apenas de empresário ou agricultor mantém-se apenas o primeiro Agente. Ademais os diálogos foram reorganizados para que todo trecho

³ *Discord* é uma ferramenta de software usado comumente por jogadores online (*gamers*) para se comunicarem ao longo do jogo. A plataforma permite a criação de servidores privados e salas de comunicação por texto e voz.

```

EmpMáq/VereAtlantisYesterday at 9:38 AM
Bom dia prezados clientes, aguardo vocês :smile:
Fiscal Atlantis/AgAT1Yesterday at 9:58 AM
Bom dia gostaria de 6 pacotes do tipo 2
EmpMáq/VereAtlantisYesterday at 9:58 AM
Bom dia @Fiscal Atlantis/AgAT1
sim, essa quantidade pelo preço normal
o que achas?
Fiscal Atlantis/AgAT1Yesterday at 9:59 AM
estou comprando em uma quantidade alta, acredito que poderia fazer pelo preço baixo
EmpMáq/VereAtlantisYesterday at 9:59 AM
vou verificar aqui, só um minuto
@Fiscal Atlantis/AgAT1 certo, faço os 6 pelo preço baixo
Fiscal Atlantis/AgAT1Yesterday at 10:00 AM
Obrigada Emp
:smile:
EmpMáq/VereAtlantisYesterday at 10:01 AM
@Fiscal Atlantis/AgAT1 pedido confirmado, obrigada pela preferência :slight_smile:

```

Figura 4 – Exemplo de um diálogo original na primeira rodada de uma das partidas do GORIM. Fonte: Própria.

de conversa apresentasse previamente o agente ao qual pertence a fala. Na Figura 5 é apresentado o mesmo trecho que está na Figura 4 entre o empresário de máquinas e o agricultor de Atlantis 1 após a reorganização, onde entende-se que cada linha apresentada contém o padrão de informação Agente Etapa 1|Agente Etapa 2 Fala Diálogo Agente. O Apêndice B contém um registro integral dos diálogos entre o empresário de máquinas e todos os agricultores pré-processado com quem realizou negociações na primeira rodada.

Posteriormente cada negociação entre o empresário em questão e cada agricultor é isolada em uma única conversa. No exemplo demonstrado na Figura 5 o diálogo já está isolado das demais como em sua versão final, compondo o conjunto dos diálogos das transações entre os agentes.

```

EmpMáq|VereadorAT      Bom dia prezados clientes, aguardo vocês :smile:
AgAT1|FiscalAT Bom dia gostaria de 6 pacotes do tipo 2
EmpMáq|VereadorAT      Bom dia @AgAT1|FiscalAT
EmpMáq|VereadorAT      sim, essa quantidade pelo preço normal
EmpMáq|VereadorAT      o que achas?
AgAT1|FiscalAT estou comprando em uma quantidade alta, acredito que poderia fazer pelo preço baixo
EmpMáq|VereadorAT      vou verificar aqui, só um minuto
EmpMáq|VereadorAT      @AgAT1|FiscalAT certo, faço os 6 pelo preço baixo
AgAT1|FiscalAT Obrigada Emp
AgAT1|FiscalAT :smile:
EmpMáq|VereadorAT      @AgAT1|FiscalAT pedido confirmado, obrigada pela preferência :slight_smile:

```

Figura 5 – Exemplo de uma transação isolada das demais e padronizada. Fonte: Própria.

3.2 Léxicos de Sentimentos

Neste trabalho os léxicos de sentimento (LS) construídos foram desenvolvidos com base nas teorias das emoções apresentadas na Seção 2.5. Deste modo, os léxicos de sentimentos foram construídos através de uma busca em profundidade, que buscou visitar até 4 níveis de profundidade a partir de um conjunto de palavras em um dicio-

nário de sinônimos online⁴. Este recurso foi construído de forma balanceada, ou seja, garante a mesma quantidade de termos para cada sentimento inicial. Este processo de construção dos léxicos de sentimentos pode ser reproduzido para a produção de outros léxicos de sentimento provenientes de outras teorias das emoções.

A partir de um conjunto de sentimentos para cada sentimento ocorre o mesmo processo onde-se para cada termo são buscados seus respectivos sinônimos. O processo de busca para uma emoção é exemplificado de maneira resumida através da Figura 6. No exemplo, o sentimento alegria é expandido em seus sinônimos em um primeiro nível, os sinônimos encontrados são posteriormente expandidos em seus respectivos sinônimos em um segundo nível, ocorrendo até o quarto nível.

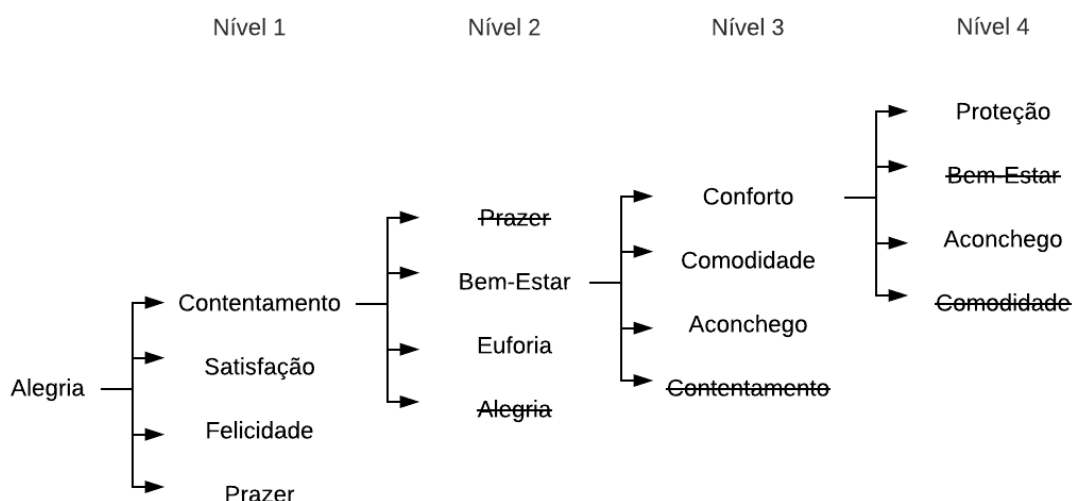


Figura 6 – Exemplo da busca em níveis do termo Alegria. Fonte: Própria

A cada expansão é verificado o balanceamento entre os conjuntos palavras para cada sentimento, caso não seja possível garantir o balanceamento entre os termos com expansões dos demais sentimentos a busca é então encerrada. Os termos tachados representam sentimentos que já teriam sido processados e não são expandidos novamente para evitar expansões repetidas.

O léxico de sentimento produzido a partir da teoria das emoções básicas de Ekman (EKMAN, 1993), foi construído a partir da busca dos termos: 1) alegria, 2) medo, 3) nojo, 4) raiva, 5) surpresa e 6) tristeza. Já o léxico de sentimento gerado a partir da roda da emoção de Plutchik (PLUTCHIK, 1991) teve sua construção a partir da busca dos termos: 1) alegria, 2) antecipação, 3) confiança, 4) medo, 5) nojo, 6) raiva, 7) surpresa e 8) tristeza.

Como resultado final tem-se dois recursos léxicos distintos com a associação de sinônimos a cada um dos sentimentos listados inicialmente. O recurso balanceado

⁴<https://www.sinonimos.com.br/>

produzido com base na teoria de Ekman tem 20.526 palavras, distribuídas em 3.421 palavras por emoção. Já o recurso balanceado produzido com base na teoria de Plutchik tem 18.864 palavras, distribuídas em 2.358 palavras por emoção.

3.3 Dicionário de Termos Intensificadores

Neste trabalho também foi construído um dicionário de termos intensificadores, com o intuito de servir como um dicionário auxiliar para a AS. Os termos foram selecionados com base em listas de advérbios de intensidade, posteriormente feito uma busca pelos seus sinônimos e determinado pesos de acordo com a intensidade na expressão para cada um. Assim os advérbios que apresentam maior intensidade em sua expressão tiveram o peso máximo, já os advérbios com menor intensidade receberam o menor peso. Por fim, aos termos intermediários foi atribuído um peso intermediário. O dicionário foi incluído integralmente no Apêndice C. O padrão de notação utilizado é de ‘termo;peso atribuído’.

3.4 Classificador de Sentimentos

O classificador de sentimentos desenvolvido neste estudo foi produzido considerando a utilização de uma abordagem baseada em léxico de sentimento, conforme explicado na Seção 2.4. Devido a dificuldade de se encontrar na literatura um léxico de sentimento disponível que associasse termos (palavras) provenientes da teoria das emoções básicas de Ekman e da roda das emoções de Plutchik, foi realizada a construção dos mesmos. Estes léxicos de sentimentos foram apresentados na Seção 3.2.

Kim; Klinger (2018b) explicam que uma metodologia diferente e conceitos diferentes ao se lidar com emoções induz cada autor para uma definição particular. Assim, definir o conceito de emoção é uma tarefa desafiadora. De acordo com Ortony; Clore; Collins (1990), teorias acerca das emoções se baseiam em todos os tipos de variáveis, não faltando tentativas de se caracterizar a estrutura das emoções.

Neste trabalho foi desenvolvido um classificador de sentimentos baseado em léxico de sentimento, com uma granularidade ao nível de sentença (ver definições na Seção 2.4), entendendo por sentença qualquer trecho de comunicação entre agentes. A classificação por sentença ocorre antes da classificação por transação, entendendo por transação qualquer ação realizada por um agente no jogo.

O classificador de sentimento faz uso da biblioteca *spaCy*⁵ que permite o reconhecimento de termos por sua classe gramatical. Ademais o classificador utiliza o dicionário de termos intensificadores, descrito na Seção 3.3. Assim, inicialmente o classificador percorre as sentenças à procura destes termos e, no caso da identifica-

⁵<https://spacy.io/>

ção de termos, o mesmo é marcado na sentença. Então, novamente a sentença é avaliada buscando-se pelo reconhecimento de termos presentes no léxico de sentimento, se algum termo é reconhecido, o mesmo também é marcado na sentença. Após, verifica-se gramaticalmente a relação entre os termos, na possibilidade do termo intensificador estar intensificando o termo proveniente do léxico de sentimento, então atribui-se o peso previamente definido, positivamente ou negativamente dependendo da emoção em questão.

Nas Figuras 7 e 8 pode-se ver o modo como a classificação é feita em cada sentença e, também, a utilização do léxico de sentimento com base nas emoções de Ekman e de Plutchik.

```

bom dia, gostaria de alugar 3 pacotes de máquinas do tipo 3
{'alegria': 1, 'medo': 1, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 1}
Bom dia... Essa quantia sai pelo valor alto... Ok?
{'alegria': 1, 'medo': 2, 'nojo': 2, 'raiva': 0, 'surpresa': 0, 'tristeza': 1}
você tem alguma promoção?
{'alegria': 0, 'medo': 0, 'nojo': 0, 'raiva': 2, 'surpresa': 0, 'tristeza': 0}
Sim... Pra compras acima de 5 unidades
{'alegria': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
:slight_smile:
{'alegria': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
o preço é baixo?
{'alegria': 2, 'medo': 0, 'nojo': 0, 'raiva': 1, 'surpresa': 0, 'tristeza': 0}
Neste caso sim
{'alegria': 3, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 3, 'tristeza': 0}
então pode ser 6
{'alegria': 1, 'medo': 1, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
obrigada
{'alegria': 0, 'medo': 1, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
Ok... 6 unidades do pacote tipo 3 pelo preço baixo
{'alegria': 1, 'medo': 0, 'nojo': 1, 'raiva': 1, 'surpresa': 0, 'tristeza': 2}
obrigado
{'alegria': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 1, 'tristeza': 0}
:slight_smile:
{'alegria': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}

```

Figura 7 – Exemplo da classificação por sentença – LS Ekman. Fonte: Própria

```

bom dia, gostaria de alugar 3 pacotes de máquinas do tipo 3
{'alegria': 1, 'antecipação': 0, 'confiança': 0, 'medo': 1, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 1}
Bom dia... Essa quantia sai pelo valor alto... Ok?
{'alegria': 1, 'antecipação': 0, 'confiança': 1, 'medo': 2, 'nojo': 1, 'raiva': 0, 'surpresa': 0, 'tristeza': 1}
você tem alguma promoção?
{'alegria': 0, 'antecipação': 1, 'confiança': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 1, 'tristeza': 0}
Sim... Pra compras acima de 5 unidades
{'alegria': 0, 'antecipação': 0, 'confiança': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
:slight_smile:
{'alegria': 0, 'antecipação': 0, 'confiança': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
o preço é baixo?
{'alegria': 1, 'antecipação': 0, 'confiança': 1, 'medo': 0, 'nojo': 0, 'raiva': 1, 'surpresa': 0, 'tristeza': 0}
Neste caso sim
{'alegria': 3, 'antecipação': 0, 'confiança': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 3, 'tristeza': 0}
então pode ser 6
{'alegria': 1, 'antecipação': 0, 'confiança': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
obrigada
{'alegria': 0, 'antecipação': 0, 'confiança': 0, 'medo': 1, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
Ok... 6 unidades do pacote tipo 3 pelo preço baixo
{'alegria': 1, 'antecipação': 0, 'confiança': 1, 'medo': 0, 'nojo': 0, 'raiva': 1, 'surpresa': 0, 'tristeza': 2}
obrigado
{'alegria': 1, 'antecipação': 0, 'confiança': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}
:slight_smile:
{'alegria': 0, 'antecipação': 0, 'confiança': 0, 'medo': 0, 'nojo': 0, 'raiva': 0, 'surpresa': 0, 'tristeza': 0}

```

Figura 8 – Exemplo da classificação por sentença – LS Plutchik Fonte: Própria

Por fim, deseja-se obter para este estudo um classificador por transação. Considerando o classificador desenvolvido por sentença capaz de, a partir de uma sentença, classificá-la em uma emoção, este classificador analisa as sentenças de uma mesma transação e define a emoção predominante da transação como sendo a emoção que mais se destaca na transação.

Nesta etapa ocorre a produção de um *dataset* intermediário das transações com sentimentos rotulados. O *dataset* rotula cada transação já ocorrida em cada jogo de acordo com o registro do diálogo das transações a um sentimento correspondente (conforme descrito na Seção 3.1.2). Nos dados provenientes do motor do RPG GO-RIM já padronizados conforme apresentado na Tabela 3 são adicionadas as informações a respeito da classificação de sentimentos das transações, assim ocorre uma integração entre os *datasets*.

Realizada esta integração entre *datasets* obtém-se o *dataset* final que será utilizado na etapa seguinte. As descrições dessas etapas realizadas podem ser vistas na Figura 9.

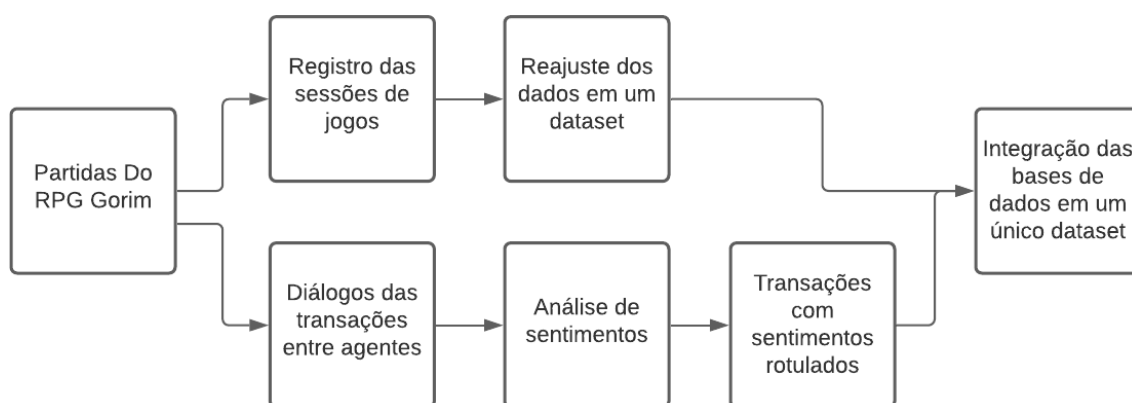


Figura 9 – Diagrama das etapas realizadas para união dos *datasets*. Fonte: Própria

3.5 Preditores

A proposta foi construir modelos de AM a partir de dados de entrada, formando um conjunto de treinamento e utilizar isto para a previsão de novos dados nunca antes vistos, como se refere Muller; Guido (2017). Assim espera-se dos algoritmos a capacidade de aprender sobre um conjunto de dados pré-definido, gerando modelos onde novas entradas possam ser analisadas e suas saídas determinadas corretamente.

Foram utilizados quatro diferentes algoritmos de AM, todos no contexto de aprendizado supervisionado, conforme definido na Seção 2.3. Os modelos produzidos durante este estudo tratam-se de modelos de regressão. Esse tipo de modelo é utilizado quando a solução do conjunto de dados para qual o algoritmo que é treinado engloba dados contínuos, ou seja, não são dados categóricos, tratando-se de um mapeamento para uma função contínua.

Todos os modelos foram desenvolvidos com utilização da biblioteca de AM *Scikit-Learn*⁶, de código aberto, disponível para uso em programação *Python*. Utiliza-se o *dataset* já integrado ao classificador de sentimentos, conforme descrito na Seção 3.4 e demonstrado através da Figura 9. A aplicação dos modelos e as etapas anteriores e posteriores estão sumarizadas na Figura 10.

Na etapa de pré-processamento eliminam-se alguns atributos do *dataset* pois ao utilizar todas as *features* inicialmente estabelecidas notou-se uma co-relação entre os atributos e a saída desejada pelos modelos. Utilizou-se de um algoritmo auxiliar de seleção de atributos para verificar a influência de cada *feature* na saída do algoritmo. A Eliminação de Atributos Recursivamente (em inglês, *Recursive Feature Elimination* – RFE) é um algoritmo disponível também na biblioteca do *Scikit-Learn*, que seleciona atributos recursivamente considerando cada vez um conjunto menor de atributos.

No *dataset* em questão foram eliminados os atributos referentes aos saldos de

⁶<https://scikit-learn.org/>

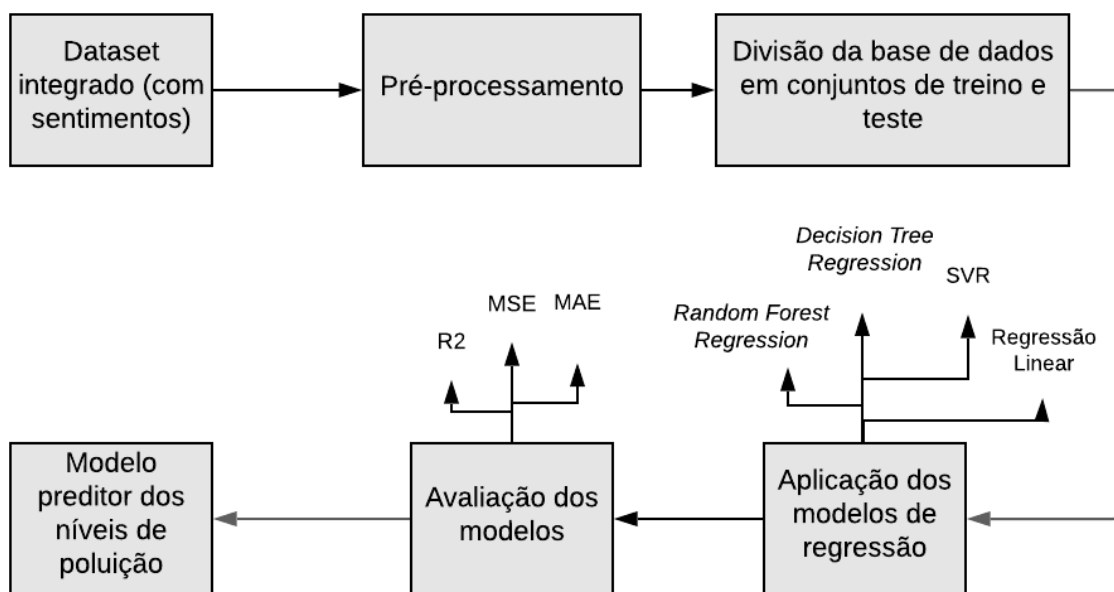


Figura 10 – Diagrama das etapas realizadas para o desenvolvimento dos modelos de AM.
Fonte: Própria.

todos os agentes, os atributos que registram os agentes envolvidos na transação, o atributo referente a parcela de terra em que a ação é realizada e o valor numérico que representa a contagem do jogo.

Após a eliminação das *features* citadas, o *dataset* teve seus valores normalizados, os dados categóricos são transformados em indicadores *booleanos* e os valores contínuos são padronizados em um intervalo de $[0,1]$, para então o conjunto de dados ser dividido em conjuntos de treino e teste.

4 RESULTADOS

Neste Capítulo são apresentados os resultados obtidos por cada algoritmo de regressão utilizando-se das métricas de avaliação: R^2 , MAE e o MSE, conforme apresentadas na Seção 2.3, considerando dois cenários. São eles: Cenário 1 utilização do classificador de sentimento baseado no léxico de sentimento construído levando em consideração a teoria das emoções básicas de Ekman e Cenário 2 utilização do classificador de sentimento baseado no léxico de sentimento construído levando em consideração a roda as emoções de Plutchik.

Com o propósito de se destacar as técnicas utilizadas em relação ao que se encontra na literatura e sumarizar isto, apresenta-se novamente a Tabela 4 (inicialmente apresentada na Seção 2.6) para que seja possível observar as diferenças entre a AS nesta dissertação e nos trabalhos relacionados.

Tabela 4 – Comparação entre os trabalhos relacionados e a abordagem proposta.

Trabalho	Algoritmos de Classificação	Ferramenta	Datasets	Léxicos
(KESHTKAR et al., 2012)	Naive Bayes (NB), Máquina de Vetores Suporte (SVM) e Árvore de Decisão (J48)	Weka	Chat de conversa do jogo <i>Land Science</i>	SentiWordNet, Subjectivity lexicon, General Inquirer, Taboada e LIWC
(KANSAON; BRANDÃO; PINTO, 2018)	Algoritmos Baseados no Modelo <i>Naive Bayes</i>	Weka	Twitter	Não utiliza
(GAIND; SYAL; PADGALWAR, 2019)	Otimização Mínima Sequencial (SMO) e Árvore de Decisão (J48)	Weka	Twitter	<i>Emotion-Words Set</i> (EWS), <i>Degree-Words Set</i> (DWS) e <i>Location-Areas Set</i> (LAS)
Abordagem Proposta	Não utiliza	Scikit-Learn	Chat de conversas das negociações do jogo RPG GORIM	LS Cenário 1 - Ekamn, LS Cenário 2 - Plutchik e Dicionário de termos intensificadores

4.1 Cenário 1

Esta Seção aborda os resultados dos modelos de regressão (Seção 3.5) construídos para a predição dos níveis de poluição do ambiente do RPG GORIM. Este Cenário 1 utiliza-se além dos dados provenientes do motor do RPG GORIM, as transações dos agentes classificadas em sentimentos. A classificação das emoções neste Cenário é baseada na teoria das emoções básicas de Ekman (Seção 3.2) através do uso do lé-

xico de sentimento produzido. Os resultados da avaliação das métricas utilizadas para avaliação dos modelos de regressão podem ser vistos na Tabela 5.

Tabela 5 – Cenário 1: Avaliação dos modelos.

Modelo	MSE	MAE	R^2
Regressão Linear	0.018	0.108	0.647
<i>Decision Tree Regression</i>	0.013	0.085	0.740
<i>Random Forest Regression</i>	0.013	0.086	0.740
<i>Support Vector Regression</i>	0.015	0.097	0.708

Observando-se os valores apresentados pelos modelos, para cada métrica, o modelo *Decision Tree Regression* é ligeiramente melhor que o modelo *Random Forest Regression*. Considerando que ambos, ainda que de modo diferente, tenham como base em seu processo de regressão árvores de decisão é esperada alguma similaridade nos resultados desses modelos. O *Support Vector Regression* tem um MAE e MSE maior do que os modelos anteriores, porém seu R^2 é menor do que os modelos anteriores. A Regressão Linear em comparação ao *Decision Tree Regression* e *Random Forest Regression* possuiu assim como o SVR um MAE e MSE um pouco maior do que os modelos anteriores e um R^2 um pouco menor do que os modelos anteriores.

Através de uma amostra da população (parte do conjunto de dados utilizado para teste) foram gerados os gráficos abaixo, possibilitando a visualização entre os níveis gerais produzidos de poluição e os valores preditos por cada modelo. Os níveis de poluição são apresentados no eixo y do gráfico e o eixo x não demonstra relação temporal. Os gráficos da Regressão Linear, *Decision Tree Regression*, *Random Forest Regression* e *Support Vector Regression* podem ser vistos respectivamente nas Figuras 11, 12, 13 e 14.

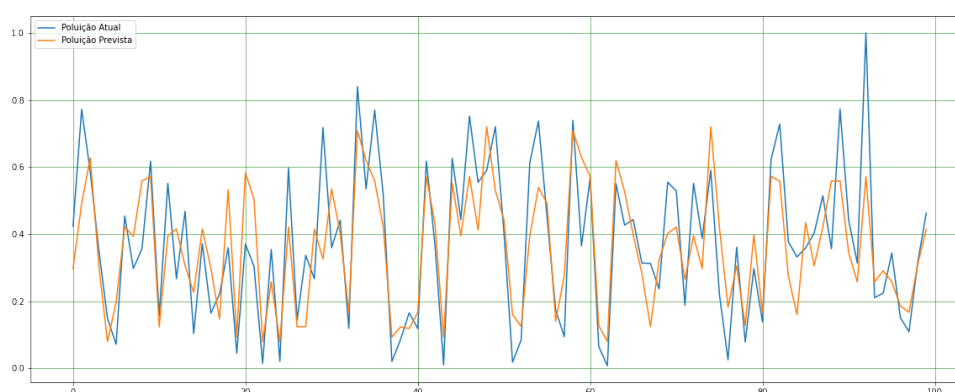


Figura 11 – Cenário 1: Amostra do Modelo de Regressão Linear. Fonte: Própria.

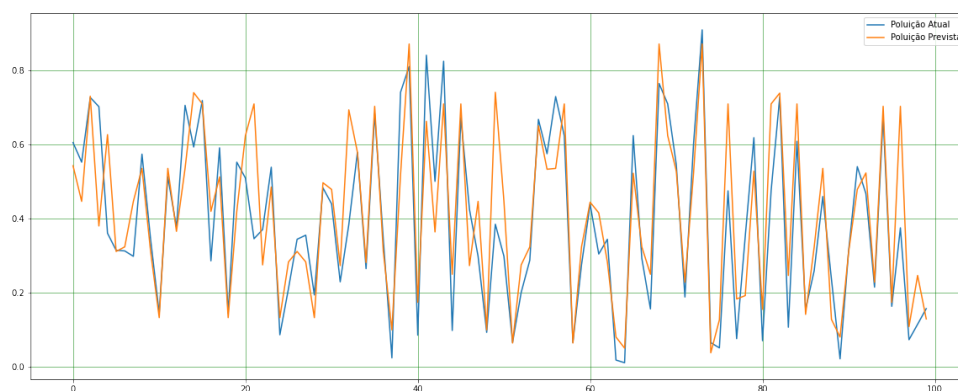


Figura 12 – Cenário 1: Amostra do Modelo de *Decision Tree Regression*. Fonte: Própria.

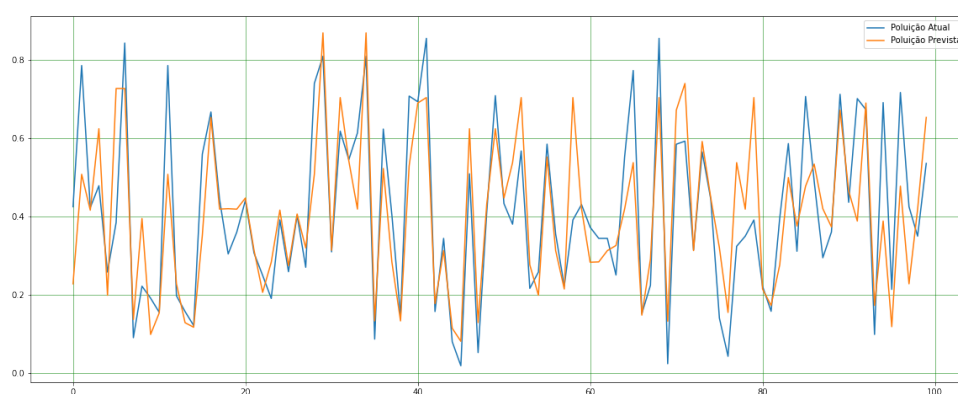


Figura 13 – Cenário 1: Amostra do Modelo de *Random Forest Regression*. Fonte: Própria.

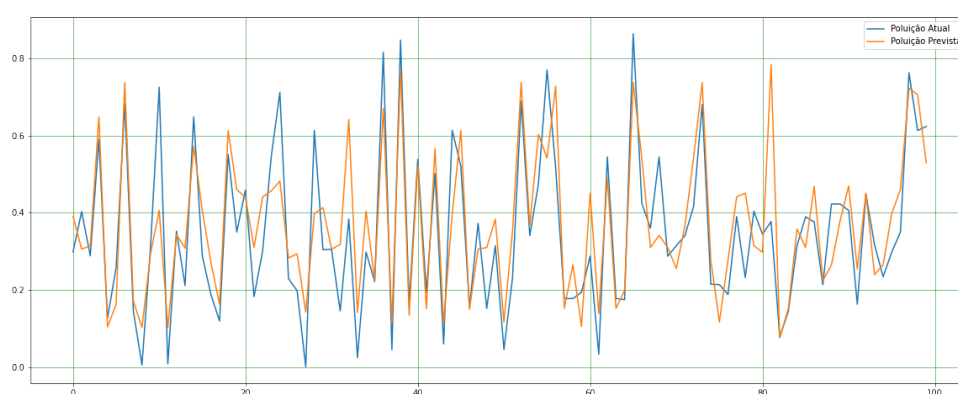


Figura 14 – Cenário 1: Amostra do Modelo de *Support Vector Regression*. Fonte: Própria.

4.2 Cenário 2

Esta Seção aborda os resultados dos modelos de regressão (Seção 3.5) em um segundo cenário. A classificação das emoções neste Cenário é baseada na teoria

das emoções básicas de Plutchik (Seção 3.2) através do uso do léxico de sentimento produzido. Os resultados da avaliação das métricas utilizadas para avaliação dos modelos de regressão podem ser vistos na Tabela 6.

Tabela 6 – Cenário 2: Avaliação dos modelos.

Modelo	MSE	MAE	R^2
Regressão Linear	0.017	0.107	0.652
<i>Decision Tree Regression</i>	0.013	0.085	0.737
<i>Random Forest Regression</i>	0.013	0.086	0.742
<i>Support Vector Regression</i>	0.015	0.097	0.705

Observando-se os valores apresentados pelos modelos para cada métrica, os modelos *Decision Tree Regression* e *Random Forest Regression* possuem valores próximos, de modo que o *Decision Tree Regression* tem melhores valores para MSE e MAE e o *Random Forest Regression* tem melhor valor de R^2 . Como discutido no Cenário 1 (Seção 4.1), ambos algoritmos de regressão utilizam em seu processo árvores de decisão. O *Support Vector Regression* se aproxima mais dos valores preditos pela Regressão Linear. Em comparação aos dois modelos anteriores a Regressão Linear e o SVR, possuem um MAE e MSE ligeiramente maior do que os anteriores e um R^2 menor do que os anteriores.

Através de uma amostra da população (parte do conjunto de dados utilizado para teste) foram gerados os gráficos abaixo, possibilitando a visualização entre os níveis gerais produzidos de poluição e os valores preditos por cada modelo. Os níveis de poluição são apresentados no eixo y do gráfico e o eixo x não demonstra relação temporal. Os gráficos da Regressão Linear, *Decision Tree Regression*, *Random Forest Regression* e *Support Vector Regression* podem ser vistos respectivamente nas Figuras 15, 16, 17 e 18.

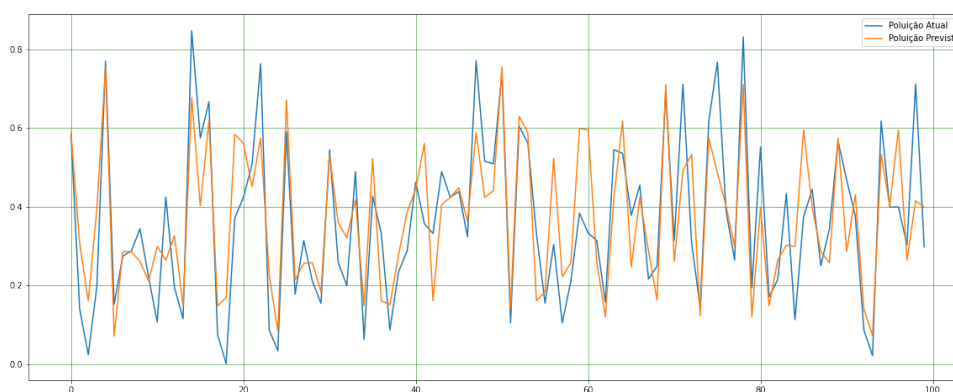


Figura 15 – Cenário 2: Amostra do Modelo de Regressão Linear. Fonte: Própria.

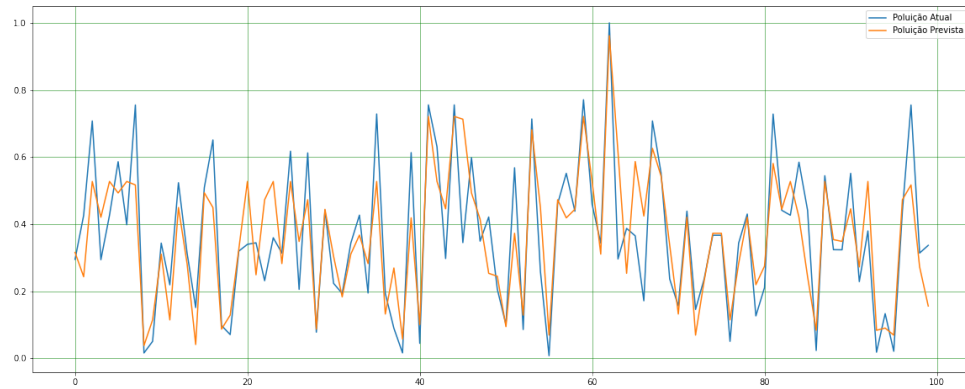


Figura 16 – Cenário 2: Amostra do Modelo de *Decision Tree Regression*. Fonte: Própria.

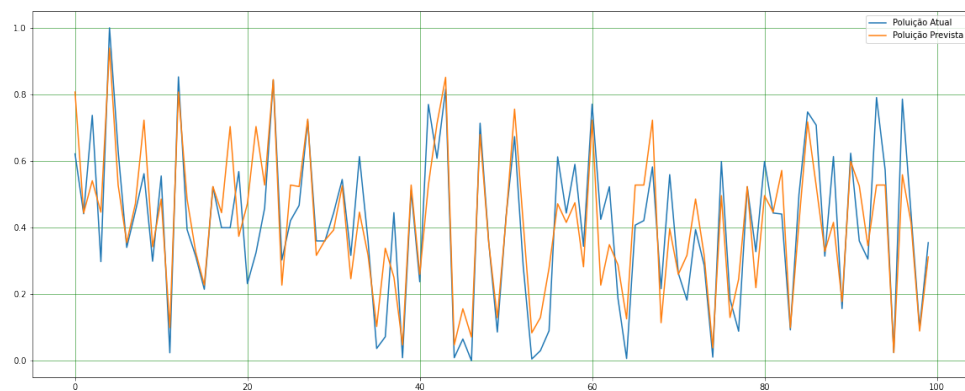


Figura 17 – Cenário 2: Amostra do Modelo de *Random Forest Regression*. Fonte: Própria.

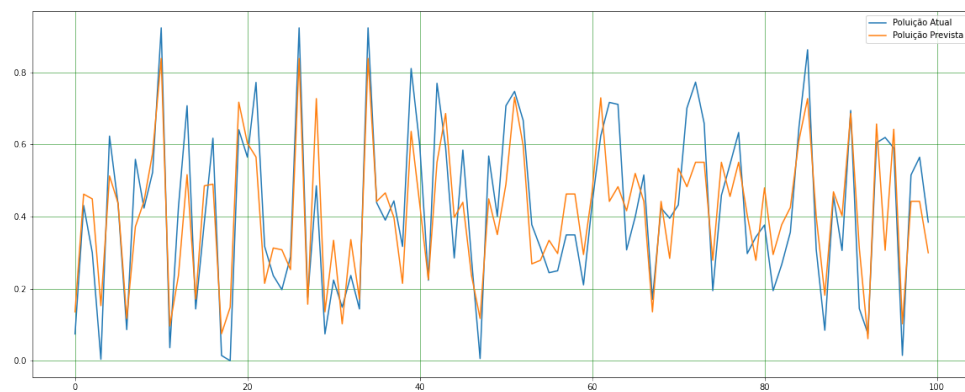


Figura 18 – Cenário 2: Amostra do Modelo de *Support Vector Regression*. Fonte: Própria.

5 CONCLUSÕES

O manejo dos recursos hídricos deve ser feito racionalmente, garantindo o desenvolvimento e a preservação do meio ambiente. Um bom gerenciamento e aproveitamento dos recursos hídricos auxilia no combate dos efeitos adversos da poluição, da seca e do assoreamento. Ademais, uma boa gestão implica em um uso hídrico consciência em função dos seus diferentes tipos de uso, das diferentes áreas de conhecimento envolvidas.

A integração de recursos e tecnologias torna possível compartilhar o conhecimento de outros modos. No contexto do grupo de pesquisa da qual este trabalho se insere está sendo desenvolvido um jogo de RPG para a gestão participativa dos recursos hídricos que é amplamente utilizado como ferramenta de coleta de dados nesta dissertação.

Neste jogo os níveis de poluição ambiente jogo são utilizados para se medir quão saudável está o meio ambiente. A partir das partidas do RPG GORIM são gerados dois tipos de dados diferentes. Os diálogos entre os personagens fazendo as transações conforme descrito na Seção 3.1.2 e os dados provenientes do motor RPG GORIM que são tratados conforme descrito na Seção 3.1.1 que refletem as decisões tomadas nas negociações.

O objetivo desta dissertação consiste na construção de modelos de predição dos níveis de poluição com o uso de atributos subjetivos, onde por atributos subjetivos compreende-se a utilização das técnicas de AS para a classificação das transações realizadas pelos jogadores.

A partir deste objetivo, diferentes algoritmos de regressão foram abordados e aplicados neste estudo, mesmo com leve diferenças ainda sim todos atingiram bons resultados ao se analisar as métricas apresentadas no Capítulo 4. Os mesmos algoritmos de regressão foram aplicados considerando-se as diferentes teorias de emoções abordadas por esta dissertação, no caso a teoria de Ekman e a roda de emoções de Plutchik.

Assim, demonstrou-se que a abordagem utilizada neste estudo permite que outro conjunto de sentimentos seja utilizado como atributo subjetivo aos modelos. Ademais,

os modelos construídos permitem que ao longo da partida do RPG GORIM através das negociações já seja possível prever os níveis de poluição já que as transações somente são concluídas ao fim de cada rodada.

O destaque desta dissertação está no classificador de sentimento baseado em léxico de sentimento, o qual pode ser utilizado com outros léxicos de sentimentos. Considerando os modelos de regressão construídos destacam-se os modelos desenvolvidos com base nos algoritmos *Decision Tree Regression* e *Random Forest Regression* apresentando os melhores resultados nas métricas de avaliação utilizadas em ambos cenários propostos (Cenário 1 e Cenário 2).

Entretanto, os demais modelos, de Regressão Linear e SVR não apresentam resultados distantes e portanto não devem ser descartados. Todo o trabalho desenvolvido, incluindo os *datasets* construídos, o dicionário de termos intensificadores, os códigos, o processo de integração dos *datasets* e os modelos estão disponibilizados na plataforma *GitHub*¹.

Como trabalhos futuros se poderia verificar a correlação do modo como os sentimentos dos jogadores evoluem e se modificam ao longo de tempo e como isto impacta nas transações e nos níveis de poluição. Ainda, neste contexto poderiam ser abordados e analisados outros modelos de predição associados à análise de sentimentos.

¹<https://github.com/karinepestana/ClassificadordeSentimentos> e <https://github.com/karinepestana/rpg-datasets-dissertacao>

REFERÊNCIAS

ADAMATTI, D. F. **Inserção de jogadores virtuais em jogos de papéis para uso em sistema de apoio a decisão em grupos**: um Experimento no Domínio da Gestão dos Recursos Naturais. 2007. Tese (Doutorado em Sistemas Digitais) — Escola Politécnica, Universidade de São Paulo, São Paulo.

ADAMATTI, D. F.; AGUIAR, M. S. de. Computational Techniques to Manage Natural Resources. **International Journal of Organizational and Collective Intelligence**, Pensilvânia, EUA, v.3, n.2, p.117–137, Apr. 2012.

ALPAYDIN, E. **Introduction to machine learning**. Massachusetts, USA: MIT press, 2020.

ARAÚJO, M.; GONÇALVES, P.; BENEVENUTO, F.; CHA, M. Métodos para análise de sentimentos no twitter. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB (WEBMEDIA'13), 19., 2013, Salvador, Bahia. **Anais...** SBC, 2013. p.19.

ASSOCIATION, A. P. **Circumplex**. (Acessado em 19/04/2021). Disponível em: <<https://dictionary.apa.org/circumplex>>.

AWAD, M.; KHANNA, R. **Support Vector Regression**. Berkeley, CA: Apress, 2015. 67–80p.

BANDINI, S.; MANZONI, S.; VIZZARI, G. RPG-Profler: a MAS for Role Playing Games Based Tests in Employee Assessment. In: WOA 2002: DAGLI OGGETTI AGLI AGENTI. 3RD AI*IA/TABOO JOINT WORKSHOP "FROM OBJECTS TO AGENTS": FROM INFORMATION TO KNOWLEDGE, 2002. **Proceedings...** Pitagora Editrice Bologna, 2002. p.72–77.

BARRETEAU, O.; PAGE, C. L.; D'AQUINO, P. Role-Playing Games, Models and Negotiation Processes. **Journal of Artificial Societies and Social Simulation**, United Kingdom, v.6, n.2, p.1–10, 2003.

BORN, M. B. **Sobre Linguagens para Modelagem de Sistemas Multiagentes em Domínio Ecológico**. UFPel: Programa de Pós-Graduação em Computação, 2018. (Exame de Qualificação).

BREIMAN, L. Random Forests. **Mach. Learn.**, USA, v.45, n.1, p.5–32, Oct. 2001.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification and Regression Trees**. Monterey, CA: Wadsworth and Brooks, 1984.

BROWNLEE, J. **What Are Word Embeddings for Text?** (Acessado em 09/03/2021). Disponível em: <<https://machinelearningmastery.com/what-are-word-embeddings/>>.

CAMBRIA, E. Affective Computing and Sentiment Analysis. **IEEE Intelligent Systems**, Piscataway, Nova Jersey, EUA, v.31, n.2, p.102–107, Mar 2016.

CARDOSO, P. **Aplicativo Discord junta gamers em chats de voz**. Disponível em: <<https://www.techtudo.com.br/tudo-sobre/discord.html>>. Acesso em: 2021-02-15.

CARVALHO, P.; SILVA, M. J. SentiLex-PT: Principais características e potencialidades. **Oslo Studies in Language**, Noruega, v.7, n.1, 2015.

DARWIN, C.; PRODGER, P. **The expression of the emotions in man and animals**. USA: Oxford University Press, 1998.

DONALDSON, M. **Plutchik's Wheel of Emotions: Feelings Wheel**. USA: 6Seconds, 2017. (Acessado em 22/04/2021). Disponível em: <<https://www.6seconds.org/2020/08/11/plutchik-wheel-emotions/>>.

EKMAN, P. Facial expression and emotion. **American psychologist**, USA, v.48, n.4, p.384, 1993.

EL NAQA, I.; MURPHY, M. J. What is machine learning? In: **Machine Learning in Radiation Oncology**. Berlim, Alemanha: Springer, 2015. p.3–11.

ESULI, A.; SEBASTIANI, F. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In: FIFTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC'06), 2006, Genoa, Italy. **Proceedings...** European Language Resources Association (ELRA), 2006.

FARIAS, G. et al. Water Resources Analysis: An Approach based on Agent-Based Modeling. **Revista de Informática Teórica e Aplicada**, Rio Grande do Sul, Brasil, v.27, n.2, p.81–95, 2020.

FARIAS, G. P. et al. Modelagem Baseada em Agentes para Análise de Recursos Hídricos. In: WORKSHOP-ESCOLA DE AGENTES, SEUS AMBIENTES E APLICAÇÕES – WESAAC, 13., 2019, Santa Catarina, Brasil. **Anais...** SBC, 2019.

FELDMAN, R. Techniques and applications for sentiment analysis. **Communications of the ACM**, NY, USA, v.56, n.4, p.82–89, 2013.

FUEYO, E. **Understanding What is Behind Sentiment Analysis – Part 2**. (Acessado em 08/03/2021). Disponível em: <<https://www.kdnuggets.com/2018/04/understanding-behind-sentiment-analysis-part-2.html>>.

GAIND, B.; SYAL, V.; PADGALWAR, S. Emotion detection and analysis on social media. **Global Journal of Engineering Science and Researches**, India, 2019.

GIGANDET, S. **Circumplex**. (Acessado em 19/04/2021). Disponível em: <<https://lotsofwords.com/definition/circumplex>>.

GILMAN, R. C. The General Inquirer: A Computer Approach to Content Analysis. **American Journal of Sociology**, USA, v.73, n.5, p.634–635, Mar. 1968.

GLEN, S. **Mean Squared Error**: Definition and Example. (Acessado em 09/03/2021). Disponível em: <<https://www.statisticshowto.com/mean-squared-error/>>.

GLEN, S. **Coefficient of Determination (R Squared)**: Definition, Calculation. (Acessado em 09/03/2021). Disponível em: <<https://www.statisticshowto.com/probability-and-statistics/coefficient-of-determination-r-squared/>>.

GLEN, S. **Absolute Error & Mean Absolute Error (MAE)**. (Acessado em 09/03/2021). Disponível em: <<https://www.statisticshowto.com/absolute-error/>>.

HARRINGTON, P. **Machine learning in action**. NY, USA: Manning Publications Co., 2012.

HOLMES, G.; DONKIN, A.; WITTEN, I. H. Weka: A machine learning workbench. In: ANZIIS'94-AUSTRALIAN NEW ZEALAND INTELLIGENT INFORMATION SYSTEMS CONFERENCE, 1994, Piscataway, Nova Jersey, EUA. **Proceedings...** IEEE, 1994. p.357–361.

HOLZMAN, B. **Natural Resource Management**. (Acessado em 30/03/2020). Disponível em: <<http://online.sfsu.edu/bholzman/courses/GE0G%20657/env%20history%20lecture.pdf>>.

HURWITZ, J.; KIRSCH, D. Machine learning for dummies. **IBM Limited Edition**, NJ, USA, v.75, 2018.

HUSSEIN, D. M. E.-D. M. A survey on sentiment analysis challenges. **Journal of King Saud University-Engineering Sciences**, Holanda, v.30, n.4, p.330–338, 2018.

KAEHLING, L. P.; LITTMAN, M. L.; MOORE, A. W. Reinforcement learning: A survey. **Journal of artificial intelligence research**, CA, USA, v.4, p.237–285, 1996.

KANSAON, D. P.; BRANDÃO, M. A.; PINTO, S. A. P. de. Análise de Sentimentos em Tweets em Português Brasileiro. In: VII BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 2018, Natal, Brasil. **Anais...** SBC, 2018.

KAVITHA, S.; VARUNA, S.; RAMYA, R. A comparative analysis on linear regression and support vector regression. In: INTERNATIONAL CONFERENCE ON GREEN ENGINEERING AND TECHNOLOGIES (IC-GET), 2016, Piscataway, Nova Jersey, EUA. **Proceedings...** IEEE, 2016. p.1–5.

KESHTKAR, F.; BURKETT, C.; GRAESSER, A.; LI, H. Detecting players personality behavior with any effort of concealment. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TEXT PROCESSING AND COMPUTATIONAL LINGUISTICS, 2012, Berlin, Alemanha. **Proceedings...** Springer, 2012. p.502–514.

KIM, E.; KLINGER, R. Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 27., 2018, New Mexico, USA. **Proceedings...** Association for Computational Linguistics, 2018. p.1345–1359.

KIM, E.; KLINGER, R. **A survey on sentiment and emotion analysis for computational literary studies**. USA: Cornell University, 2018.

LEITZKE, B. et al. Sistema multiagente para gestão de recursos hídricos: Modelagem da bacia do São Gonçalo e da Lagoa Mirim. In: X WORKSHOP DE COMPUTAÇÃO APLICADA A GESTÃO DO MEIO AMBIENTE E RECURSOS NATURAIS – WCAMA, 2019, PA, Brasil. **Anais...** SBC, 2019. p.87–96.

LIU, B. Sentiment Analysis and Subjectivity. In: INDURKHYA, N.; DAMERAU, F. J. (Ed.). **Handbook of Natural Language Processing**. Londres, Inglaterra: Chapman and Hall/CRC, 2010. p.627–666.

LUTZ, M. **Programming python**. CA, USA: O'Reilly Media, Inc., 2001.

MEJOVA, Y. Sentiment analysis: An overview. **University of Iowa, Computer Science Department**, USA, 2009.

MOHSIN, M. A.; BELTIUKOV, A. Summarizing Emotions from Text Using Plutchik's Wheel of Emotions. In: SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGIES FOR INTELLIGENT DECISION MAKING SUPPORT (ITIDS 2019), 7., 2019, Berlim, Alemanha. **Proceedings...** Atlantis Press – Springer, 2019. p.291–294.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. USA: Wiley, 2012.

MULLER, A. C.; GUIDO, S. **Introduction to machine learning with Python: a guide for data scientists**. USA: O'Reilly Media, 2017.

NEETHU, M.; RAJASREE, R. Sentiment analysis in twitter using machine learning techniques. In: IV INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATIONS AND NETWORKING TECHNOLOGIES (ICCCNT), 2013, Piscataway, Nova Jersey, EUA. **Proceedings...** IEEE, 2013. p.1–5.

NISBET, R.; ELDER, J.; MINER, G. Chapter 7 - Basic Algorithms for Data Mining: A Brief Overview. In: NISBET, R.; ELDER, J.; MINER, G. (Ed.). **Handbook of Statistical Analysis and Data Mining Applications**. Boston: Academic Press, 2009. p.121 – 150.

ORTONY, A.; CLORE, G. L.; COLLINS, A. **The cognitive structure of emotions**. Inglaterra: Cambridge university press, 1990.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment classification using machine learning techniques. **Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)**, USA, 2002.

PENNEBAKER, J. W.; BOOTH, R. J.; FRANCIS, M. E. Linguistic inquiry and word count: LIWC [Computer software]. **LIWC.net**, Texas, USA, v.135, 2007.

PERROTON, A.; GARINE-WICHATITSKY, M. de; VALLS-FOX, H.; LE PAGE, C. My cattle and your park: codesigning a role-playing game with rural communities to promote multistakeholder dialogue at the edge of protected areas. **Ecology and Society**, NY, USA, v.22, n.1, 2017.

PLUTCHIK, R. **The emotions**. USA: University Press of America, 1991.

PONTE, B.; FUENTE, D. De la; PARREÑO, J.; PINO, R. Intelligent decision support system for real-time water demand management. **International Journal of Computational Intelligence Systems**, Berlim, Alemanha, v.9, n.1, p.168–183, 2016.

PORTAL EDUCAÇÃO. **Gestão de Recursos Hídricos - Portal Educação**. (Acessado em 04/03/2021). Disponível em: <<https://siteantigo.portaleducacao.com.br/conteudo/artigos/biologia/gestao-de-recursos-hidricos/5703>>.

SEGAL, M. Machine Learning Benchmarks and Random Forest Regression. **Technical Report, Center for Bioinformatics & Molecular Biostatistics, University of California, San Francisco, USA, 05 2003.**

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning:** From theory to algorithms. England: Cambridge university press, 2014.

STAPLETON, A. J. Serious games: Serious opportunities. In: AUSTRALIAN GAME DEVELOPERS CONFERENCE, ACADEMIC SUMMIT, 2004, Melbourne, Australia. **Proceedings...** AGDC, 2004.

SUSI, T.; JOHANNESSON, M.; BACKLUND, P. **Serious games:** An overview. Lund, Suécia: Institutionen för kommunikation och information, 2007.

TABOADA, M.; GRIEVE, J. Analyzing appraisal automatically. In: AAAI SPRING SYMPOSIUM ON EXPLORING ATTITUDE AND AFFECT IN TEXT, 2004, Palo Alto, California. **Proceedings...** Stanford University: AAAI Press, 2004. p.158–161.

TROMP, E.; PECHENIZKIY, M. **Rule-based emotion detection on social media:** putting tweets on Plutchik's wheel. USA: Cornell University, 2014.

WIEBE, J.; WILSON, T. Learning to Disambiguate Potentially Subjective Expressions. In: CONFERENCE ON NATURAL LANGUAGE LEARNING, 6., 2002, USA. **Proceedings...** Association for Computational Linguistics, 2002. p.1–7. (COLING-02, v.20).

XU, M.; WATANACHATURAPORN, P.; VARSHNEY, P. K.; ARORA, M. K. Decision tree regression for soft classification of remote sensing data. **Remote Sensing of Environment**, Holanda, v.97, n.3, p.322–336, 2005.

ZHU, X.; GOLDBERG, A. B. Introduction to semi-supervised learning. **Synthesis lectures on artificial intelligence and machine learning**, CA, USA, v.3, n.1, p.1–130, 2009.

Apêndices

APÊNDICE A – Diálogos originais retirados do *Discord* entre empresário de máquinas e agricultores (Jogo 7 - Rodada 1)

EmpMáq/VereAtlantisYesterday at 9:38 AM

Bom dia prezados clientes, aguardo vocês :smile:

AgCD3Yesterday at 9:40 AM

Bom dia

Agricultor AT2Yesterday at 9:40 AM

Bom dia! Gostaria de 3 Pacotes 1 de Máquina, por favor

AgCD3Yesterday at 9:41 AM

Gostaria de 1 pacote 1

EmpMáq/VereAtlantisYesterday at 9:41 AM

Bom dia @Agricultor AT2 como estás alugando mais pacotes posso fazer pelo preço normal

Bom dia @AgCD3 1 pacote sai pelo preço alto, se encomendar mais posso fazer o preço normal

Agricultor AT2Yesterday at 9:42 AM

@EmpMáq/VereAtlantis Ah, que ótimo! Aceito

EmpMáq/VereAtlantisYesterday at 9:42 AM

@Agricultor AT2 pedido confirmado, obrigada pela preferência :smile:

AgCD3Yesterday at 9:43 AM

@EmpMáq/VereAtlantis a partir de quantas você baixa o preço?

EmpMáq/VereAtlantisYesterday at 9:43 AM

@AgCD3 a partir de 2 já sai pelo preço normal

AgCD3Yesterday at 9:44 AM

@EmpMáq/VereAtlantis se eu levar um pacote 1 e um pacote 3, quanto me faz?

EmpMáq/VereAtlantisYesterday at 9:45 AM

@AgCD3 neste caso faço os dois pelo preço normal

AgCD3Yesterday at 9:45 AM

@EmpMáq/VereAtlantis certo, irei levar

EmpMáq/VereAtlantisYesterday at 9:46 AM

@AgCD3 pedido confirmado, obrigada pela preferência :slight_smile:

Fiscal Atlantis/AgAT1Yesterday at 9:58 AM

Bom dia gostaria de 6 pacotes do tipo 2

EmpMáq/VereAtlantisYesterday at 9:58 AM

Bom dia @Fiscal Atlantis/AgAT1

sim, essa quantidade pelo preço normal

o que achas?

Fiscal Atlantis/AgAT1Yesterday at 9:59 AM

estou comprando em uma quantidade alta, acredito que poderia fazer pelo preço baixo

EmpMáq/VereAtlantisYesterday at 9:59 AM

vou verificar aqui, só um minuto

@Fiscal Atlantis/AgAT1 certo, faço os 6 pelo preço baixo

Fiscal Atlantis/AgAT1Yesterday at 10:00 AM

Obrigada Emp

:smile:

EmpMáq/VereAtlantisYesterday at 10:01 AM

@Fiscal Atlantis/AgAT1 pedido confirmado, obrigada pela preferência :slight_smile:

AgAT3Yesterday at 10:03 AM

Bom dia, gostaria de 4 pacotes do tipo 2 e 2 pacotes do tipo 3, gostaria de preço baixo em todos

EmpMáq/VereAtlantisYesterday at 10:04 AM

Bom dia @AgAT3 essa quantidade faço pelo preço baixo

AgAT3Yesterday at 10:05 AM

fechado

EmpMáq/VereAtlantisYesterday at 10:06 AM

@AgAT3 pedido confirmado, obrigada pela preferência :slight_smile:

AgCD2|Prefeito CDYesterday at 10:14 AM

Olá Empresário, gostaria de saber como devo proceder para comprar 3 pacotes 3... Qual o valor? Você tem disponível?

AgCD3Yesterday at 10:15 AM

Bom dia @EmpMáq/VereAtlantis gostaria de comprar um pulverizador

EmpMáq/VereAtlantisYesterday at 10:15 AM

Bom dia @AgCD2|Prefeito CD gostaria de comprar 3 pacotes do tipo 3?

Tenho disponível sim, como são 3 posso fazer preço normal

AgCD3Yesterday at 10:15 AM

@EmpMáq/VereAtlantis Opa, desculpa não poderei fazer a compra, desisto

AgCD2|Prefeito CDYesterday at 10:16 AM

não, tenho interesse em colheitadeira, semeadora e drones para expandir meu negócio

EmpMáq/VereAtlantisYesterday at 10:16 AM

@AgCD3 está bem, qualquer coisa entre em contato :slight_smile:

AgCD2|Prefeito CDYesterday at 10:17 AM

O preço normal é quanto?

EmpMáq/VereAtlantisYesterday at 10:17 AM

@AgCD2|Prefeito CD sim, estas máquinas pertencem ao pacote 3

Não temos valores em números, apenas os tipos baixo, normal e alto
AgCD2|Prefeito CDYesterday at 10:18 AM
e nesse caso é baixo?
EmpMáq/VereAtlantisYesterday at 10:18 AM
Só um minuto, me perdi aqui
vou lhe informar
AgCD2|Prefeito CDYesterday at 10:19 AM
tudo bem
Aguardo para fecharmos negócio
EmpMáq/VereAtlantisYesterday at 10:19 AM
Pacote 3: baixo 85, normal 90 e alto 95
temos pulverizador também
no preço baixo 395
AgCD2|Prefeito CDYesterday at 10:20 AM
Amigo, não tenho interesse em pulverizador
Vou querer apenas 2 pacotes então
seu preço está caro
EmpMáq/VereAtlantisYesterday at 10:21 AM
Certo, de repente sua produção aumenta no próximo ano, daí fechamos
:slight_smile:
AgCD2|Prefeito CDYesterday at 10:21 AM
2 pacotes 3
EmpMáq/VereAtlantisYesterday at 10:21 AM
São os impostos
2 pacotes do tipo 3 dá pra fazer preço normal
fechamos?
AgCD2|Prefeito CDYesterday at 10:21 AM
ótimo
fechamos
EmpMáq/VereAtlantisYesterday at 10:22 AM
pedido confirmado, obrigada pela preferência :smile:

APÊNDICE B – Diálogos pré-processados entre empresário de máquinas e agricultores (Jogo 7 - Rodada 1)

EmpMaq|Vereador AT Bom dia prezados clientes, aguardo vocês :smile:

AgCD3 Bom dia

AgAT2 Bom dia! Gostaria de 3 Pacotes 1 de Máquina, por favor

AgCD3 Gostaria de 1 pacote 1

EmpMaq|Vereador AT Bom dia @AgAT2 como estás alugando mais pacotes posso fazer pelo preço normal

EmpMaq|Vereador AT Bom dia @AgCD3 1 pacote sai pelo preço alto, se encomendar mais posso fazer o preço normal

AgAT2 @EmpAgr|Vereador AT Ah, que ótimo! Aceito

EmpMaq|Vereador AT @AgAT2 pedido confirmado, obrigada pela preferência :smile:

AgCD3 @EmpAgr|Vereador AT a partir de quantas você baixa o preço?

EmpMaq|Vereador AT @AgCD3 a partir de 2 já sai pelo preço normal

AgCD3 @EmpAgr|Vereador AT se eu levar um pacote 1 e um pacote 3, quanto me faz?

EmpMaq|Vereador AT @AgCD3 neste caso faço os dois pelo preço normal

AgCD3 @EmpAgr|Vereador AT certo, irei levar

EmpMaq|Vereador AT @AgCD3 pedido confirmado, obrigada pela preferência :slight_smile:

AgAT1|Fiscal AT Bom dia gostaria de 6 pacotes do tipo 2

EmpMaq|Vereador AT Bom dia @AgAT1|Fiscal AT

EmpMaq|Vereador AT sim, essa quantidade pelo preço normal

EmpMaq|Vereador AT o que achas?

AgAT1|Fiscal AT estou comprando em uma quantidade alta, acredito que poderia fazer pelo preço baixo

EmpMaq|Vereador AT vou verificar aqui, só um minuto

EmpMaq|Vereador AT @AgAT1|Fiscal AT certo, faço os 6 pelo preço baixo

AgAT1|Fiscal AT Obrigada Emp

AgAT1|Fiscal AT :smile:

EmpMaq|Vereador AT @AgAT1|Fiscal AT pedido confirmado, obrigada pela preferência :slight_smile:

AgAT3 Bom dia, gostaria de 4 pacotes do tipo 2 e 2 pacotes do tipo 3, gostaria de preço baixo em todos

EmpMaq|Vereador AT Bom dia @AgAT3 essa quantidade faço pelo preço baixo

AgAT3 fechado

EmpMaq|Vereador AT @AgAT3 pedido confirmado, obrigada pela preferência :slight_smile:

AgCD2|Prefeito CD Olá Empresário, gostaria de saber como devo proceder para comprar 3 pacotes 3... Qual o valor? Você tem disponível?

AgCD3 Bom dia @EmpAgr|Vereador AT gostaria de comprar um pulverizador

EmpMaq|Vereador AT Bom dia @AgCD2|Prefeito CD gostaria de comprar 3 pacotes do tipo 3?

EmpMaq|Vereador AT Tenho disponível sim, como são 3 posso fazer preço normal

AgCD3 @EmpAgr|Vereador AT Opa, desculpa não poderei fazer a compra, desisto

AgCD2|Prefeito CD não, tenho interesse em colheitadeira, semeadora e drones para expandir meu negócio

EmpMaq|Vereador AT @AgCD3 está bem, qualquer coisa entre em contato :slight_smile:

AgCD2|Prefeito CD O preço normal é quanto?

EmpMaq|Vereador AT @AgCD2|Prefeito CD sim, estas máquinas pertencem ao pacote 3

EmpMaq|Vereador AT Não temos valores em números, apenas os tipos baixo, normal e alto

AgCD2|Prefeito CD e nesse caso é baixo?

EmpMaq|Vereador AT Só um minuto, me perdi aqui

EmpMaq|Vereador AT vou lhe informar

AgCD2|Prefeito CD tudo bem

AgCD2|Prefeito CD Aguardo para fecharmos negócio

EmpMaq|Vereador AT Pacote 3: baixo 85, normal 90 e alto 95

EmpMaq|Vereador AT temos pulverizador também

EmpMaq|Vereador AT no preço baixo 395

AgCD2|Prefeito CD Amigo, não tenho interesse em pulverizador

AgCD2|Prefeito CD Vou querer apenas 2 pacotes então

AgCD2|Prefeito CD seu preço está caro

EmpMaq|Vereador AT Certo, de repente sua produção aumenta no próximo ano, daí fechamos :slight_smile:

AgCD2|Prefeito CD 2 pacotes 3

EmpMaq|Vereador AT São os impostos

EmpMaq|Vereador AT 2 pacotes do tipo 3 dá pra fazer preço normal

EmpMaq|Vereador AT fechamos?

AgCD2|Prefeito CD ótimo

AgCD2|Prefeito CD fechamos

EmpMaq|Vereador AT pedido confirmado, obrigada pela preferência :smile:

APÊNDICE C – Dicionário de termos intensificadores

apenas;1
bastante;3
bem;2
certamente;3
com efeito;3
de forma nenhuma;3
de jeito nenhum;3
de modo algum;3
de muito;3
de pouco;1
de todo;3
demais;3
demasiado;3
deveras;3
efetivamente;3
em excesso;2
jamais;3
mais;2
mal;2
menos;2
muito;3
nunca;3
não;2
pior;2
pouco;1
quanto;3
quase;1
realmente;3
sim;2
talvez;1
tampouco;1
tanto;3
tão;3