

Otimização do Gerenciamento de Resíduos Sólidos Urbanos por meio de Mineração de Dados: uma Análise Empírica de Múltiplos Casos

JANAÍNA LOPES DIAS

UNIVERSIDADE DE SANTA CRUZ DO SUL (UNISC)

MICHELE KREMER SOTT

UNIVERSIDADE NOVE DE JULHO (UNINOVE)

Agradecimento à orgão de fomento:

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

OTIMIZAÇÃO DO GERENCIAMENTO DE RESÍDUOS SÓLIDOS URBANOS POR MEIO DE MINERAÇÃO DE DADOS: uma Análise Empírica de Múltiplos Casos

1 Introdução

O ritmo de crescimento da população tem influenciado a busca por melhor qualidade de vida e, conseqüentemente, aumentado o volume de resíduos gerados pelo consumo. Os Resíduos Sólidos Urbanos (RSU) são o tipo mais heterogêneo de lixo, sendo o resultado da multiplicidade de atividades causadas nos mais diferentes ambientes. A gestão inadequada de tais resíduos ocasiona impactos ambientais negativos (poluição do ar, do solo e da água), problemas de saúde (doenças causadas por insetos, roedores e diferentes formas de poluição) e segurança pública (Ferrão & Moraes, 2021; Ripa *et al.*, 2017; Yay, 2015).

De acordo com a Associação Brasileira de Empresas de Limpeza Pública e Resíduos Especiais (ABRELPE), a gestão dos resíduos sólidos tem se tornado um assunto cada vez mais complexo mundialmente ao longo da última década. No Brasil, este fato é observado devido à influência da Política Nacional dos Resíduos Sólidos (PNRS) e comprovado pelo aumento na geração de resíduos entre 2010 e 2019, que passou de 67 milhões para 79 milhões de toneladas por ano e, conseqüentemente, na geração per capita, que aumentou de 348 kg/ano para 379 kg/ano. Neste sentido, em 10 anos houve um crescimento de 19% e de 9%, respectivamente, na geração total de RSU e na geração per capita de resíduos no país (ABRELPE, 2021).

A PNRS promulgada em 2010 (Brasil, 2010) apresenta um conjunto de medidas (princípios, objetivos, instrumentos, diretrizes, metas e ações) que devem ser adotadas tendo em vista a gestão integrada e o gerenciamento adequado dos RSU, e foi elaborada tendo em vista se fazer cumprir a logística reversa. Esta lei orienta os gestores públicos, a população e as empresas, de modo que sejam aplicadas as melhores alternativas para a minimização dos impactos ambientais gerados pelo manejo inadequado dos resíduos. Além disto, a PNRS contribui com a perspectiva de uma economia circular, onde a matéria prima que foi extraída, processada e atrelada ao produto, seja inserida novamente na cadeia de produção e consumo para que os resíduos segregados possam ser novamente tratados como matérias primas, reduzindo a carga de materiais depositados em aterros e dando uma sobrevida aos mesmos.

Entre as etapas do gerenciamento de RSU, destaca-se o processo de coleta como uma das mais relevantes, por ser designada como um serviço público ou uma prestação de serviços no Brasil, e ter a sua complexidade envolvendo diferentes aspectos sociais, econômicos, políticos e ambientais (Hannan *et al.*, 2018; Guerrini *et al.*, 2017; Di Maria *et al.*, 2016). A coleta pode ser do tipo convencional (porta-a-porta) ou mecanizada, podendo ainda ter a utilização de contêineres. Neste sentido, evidencia-se a importância da segregação e do acondicionamento adequado na fonte, a fim de aumentar a valoração dos resíduos e propiciar a implantação concomitante de uma coleta seletiva (Dias *et al.*, 2021).

Segundo dados da ABRELPE (2021), no Brasil em 2010, do total de 5.565 cidades existentes, apenas 3.152 (56,6%) possuíam iniciativas relacionadas à coleta seletiva e em 2019, do total de 5.570 cidades, o número de municípios adeptos aumentou para 4.070 (73,1%). Entretanto, os índices de reciclagem nesta década permaneceram inferiores a 4% na média nacional. Apesar do avanço na adesão das cidades, a falta de políticas de incentivo para a separação dos resíduos, o baixo investimento na reciclagem e a inexistência de um mercado e de uma logística estruturada para o recebimento destes resíduos, refletem na sobrecarga do sistema de destinação final e na extração de recursos naturais, muitos já próximos do esgotamento (ABRELPE, 2021).

Em termos de coleta de RSU, a quantidade de resíduos coletados no Brasil em uma década passou de 59 milhões de toneladas em 2010 para 72,7 milhões de toneladas em 2019, e a cobertura de coleta passou de 88% para 92%, respectivamente. No estado do Rio Grande do

Sul, foco deste estudo, a coleta aumentou o seu índice de cobertura passando de 91,9% em 2010 para 95,5% em 2019. Em comparação com os demais 26 estados do Brasil, quanto ao índice de cobertura de coleta, o Rio Grande do Sul estava como sexto colocado, atrás dos estados do Amapá (97,5%), Distrito Federal (96,5%), São Paulo (95,9%), Rio de Janeiro (95,5%) e Minas Gerais (95,3%) em 2010, e como quinto colocado em 2019, atrás dos estados de São Paulo (99,6%), Rio de Janeiro (99,5%), Santa Catarina (96,2%) e Goiás (96,1%). Cabe salientar que apesar do avanço registrado, o *déficit* na abrangência deste serviço significa que 6,3 bilhões de toneladas/ano continuam sendo descartadas de maneira incorreta no meio ambiente (ABRELPE, 2021).

Com o propósito de manter índices de cobertura em ascensão, é imperativo que o dimensionamento da frota de coleta, da mão-de-obra empregada e da logística operacional sejam otimizados, considerando e respeitando as características locais, e buscando atingir a cobertura completa da população dos municípios. Para o estabelecimento dos roteiros de coleta, a metodologia utilizada em quase todas as cidades brasileiras ainda tácita, advinda dos gestores responsáveis pelos respectivos municípios (Dias et al., 2021). Também é importante salientar que a atividade demanda altos custos (Kinobe et al., 2015; Pires et al., 2017), portanto, quanto mais otimizada a rota de coleta de resíduos, maior será a redução de impactos socioeconômicos e ambientais (Hannan et al., 2018).

Conforme informações da ABRELPE (2021), no Brasil, os recursos aplicados pelos municípios na coleta de resíduos e demais serviços de limpeza urbana (destinação final de RSU, serviços de varrição, capina, limpeza e manutenção de parques jardins e outros) passaram de R\$ 17,65 bilhões (média de R\$ 8 por habitante/mês) em 2010, para R\$ 25 bilhões (R\$ 10 por habitante/mês) em 2019. O número de empregos diretos no setor (essencial e de serviço contínuo) também aumentou na respectiva década, passando de 284 mil para 332 mil empregos, crescimento de 17% dos postos de trabalho no país (ABRELPE, 2021).

Os dados acima corroboram que ao longo da primeira década da implementação da PNRS, apesar do aumento na geração de RSU, não houve um aumento proporcional de investimentos na ampliação e modernização dos serviços. É notório que mesmo sendo uma das prestações de serviço que mais requisita recursos aos cofres municipais, ainda existe carência de fundos para custear as operações do setor, visto que os valores aplicados não acompanham proporcionalmente o aumento da geração de RSU. A ABRELPE (2021) coloca ainda que nesta década a destinação final inadequada cresceu cerca de 30%, aumentando as consequências danosas para o meio ambiente e para a saúde da população.

Além disso, estima-se que a geração de RSU nas próximas décadas poderá aumentar em 50% até 2050 (alcançando 100 milhões de toneladas em 2033), em comparação ao ano de 2019 e com o crescimento populacional esperado de 12% (ABRELPE, 2021). Portanto, é importante evidenciar que novas tecnologias que auxiliem na minimização de impactos oriundos da coleta e destinação dos resíduos devem ser pesquisadas e difundidas visando a otimização dos roteiros de coletas e a destinação adequado dos RSU, a fim de minimizar os impactos ambientais (menores cargas de resíduos em aterros), reduzir os gastos de energia (menores consumos de combustíveis - óleo diesel) e contribuir para que a cobrança por estes serviços, normalmente executados pelas prefeituras municipais, possam ser mais assertivos, justos e valorados por toda a população.

Com esta preocupação em mente, nos últimos anos diversas pesquisas buscaram apresentar alternativas com modelos de otimização que auxiliem a tomada de decisão. Entre as propostas discutidas na literatura, o uso de dados coletados a partir da medição do processo tornaram-se amplamente adotadas, uma vez que visam desenvolvê-lo por meio da transformação do conhecimento tácito em explícito. Em razão disso, percebe-se um aumento na disponibilização de inventários de dados de RSU e a aplicação de ferramentas de inteligência artificial como estratégias para planejar e implementar sistemas de gestão de resíduos baseado

em previsão (Coskuner *et al.*, 2020). As técnicas de mineração de dados, por exemplo, permitem criar modelos que apoiam a definição de metodologias, tecnologias de tratamento e de locais para disposição dos RSU, bem como planejamento da capacidade de instalações de recuperação de materiais e otimização da utilização de pessoal e recursos (Ceylan, 2020). Além disso, com base nos resultados da adoção destas estratégias, os pesquisadores têm demonstrado a importância da previsão para o gerenciamento de RSU (Dai, Nie & Chen, 2020; Kontokosta *et al.*, 2018) e reforçado o valor da análise de dados para investigar os efeitos dos mais diversos fatores na quantidade e composição de geração de RSU (Ceylan, 2020).

A partir deste contexto, esta pesquisa tem como objetivo identificar padrões de coleta e a melhor rota de coleta de resíduos sólidos urbanos em três municípios do sul do Brasil (Ijuí, Itaqui e São Gabriel – localizados no estado do Rio Grande do Sul). Com este propósito, técnicas e algoritmos de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases* - KDD) e Mineração de Dados (DM) foram utilizados para analisar dados históricos como peso da coleta, rotas, dados sociodemográficos e climáticos, a fim de propor modelos e prever a geração de resíduos.

Estudos recentes têm mostrado que a chuva pode afetar a coleta, pois os resíduos podem ser transportados pela água das chuvas e depositados em locais inadequados, resultando em contaminação ambiental (Gouveia, 2012). A chuva também pode afetar o destino final dos resíduos, já que a decomposição anaeróbica da matéria orgânica no lixo gera grandes quantidades de gases de efeito estufa, principalmente metano, que contribui para o aquecimento global. Além disso, a água pode alterar significativamente o peso dos resíduos coletados. Devido a estes fatores, o volume de precipitação também é analisado neste estudo.

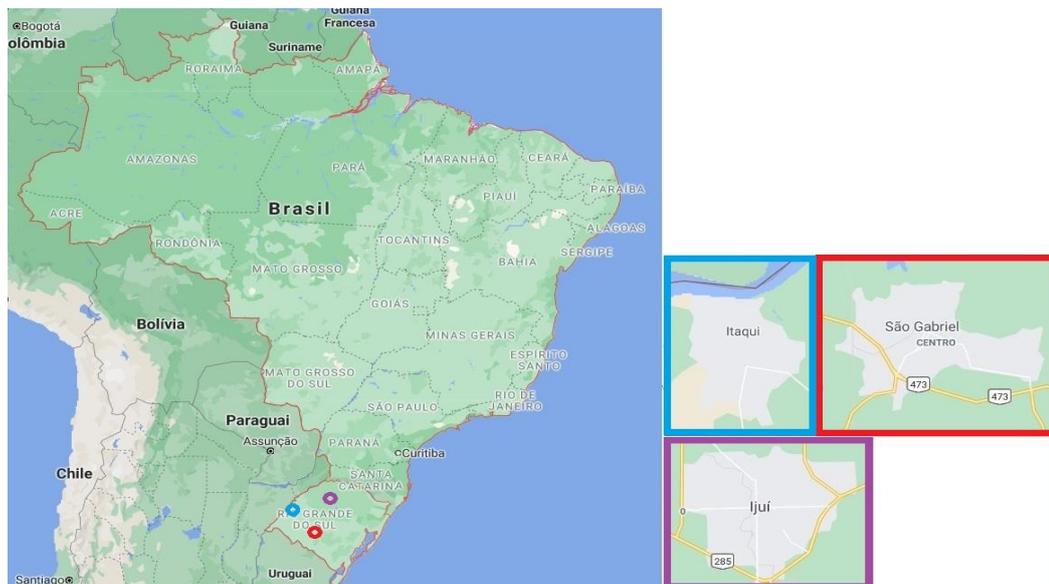
Esta pesquisa está organizada nas seguintes seções: a Seção 1 introduz o tema e objetivo do estudo, a seção 2 detalha a metodologia adotada, a seção 3 apresenta e discute os resultados e, por fim, a seção 4 contém as conclusões, limitações e sugestões de trabalhos futuros.

2 Metodologia

2.1. Área de estudo

Os modelos foram desenvolvidos a partir dos municípios de Ijuí, Itaqui e São Gabriel, localizados no Rio Grande do Sul, na região noroeste do estado, conforme Figura 1. O município de Ijuí possui uma população aproximada de 83.764 habitantes e densidade demográfica de 114,51 hab/km² (IBGE, 2021). Considerando que a coleta per capita estimada para o estado do Rio Grande do Sul no ano de 2019 foi de 264,6 kg/hab/ano e 0,72 kg/hab/dia (ABRELPE, 2021), estima-se que no município de Ijuí no ano de 2019 tenham sido coletados 22.163,95 toneladas. Do mesmo modo, Itaqui que possui uma população estimada de 37.489 habitantes e densidade demográfica de 11,21 hab/km² (IBGE, 2021) deve ter coletado aproximadamente 9.919,5 toneladas, enquanto São Gabriel, cuja população estimada em 2020 era 62.147 habitantes e densidade demográfica de 12,03 hab/km² (IBGE, 2021), deve ter coletado 16.444,09 toneladas.

Figura 1. Localização do Brasil, do estado do Rio Grande do Sul e dos municípios de Ijuí, Itaqui e São Gabriel.



Nos três municípios, os RSU são coletados por uma empresa terceirizada, por meio de contrato de prestação de serviços por período pré-determinado em processo licitatório. A coleta é feita de modo convencional (porta-a-porta), com equipes formadas por um motorista e três coletores, que recolhem os RSU domiciliares e comerciais dispostos nas vias e logradouros dos municípios. O deslocamento pelas vias públicas é realizado por caminhão basculante com equipamento compactador com capacidade de 12 ou 15 m³.

Em decorrência de não existir aterros sanitários nestes municípios, os resíduos de Ijuí são transportados para o município de Giruá (distância de 79 km), os de Itaqui são transportados para Candiota (distância de 306 km) e São Gabriel encaminha para Santa Maria (distância de 162 km). Os aterros sanitários de Giruá, Candiota e Santa Maria estão devidamente licenciados pelo órgão competente estadual para receberem os RSU.

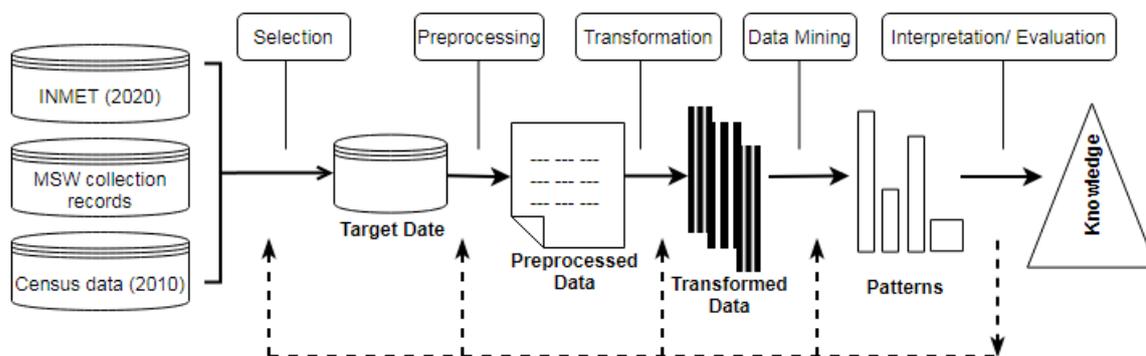
Sendo assim, os três municípios possuem no seu território unidade de transbordo de RSU, que são pavilhões adequadamente cobertos, com piso impermeabilizado, paredes em concreto, com área própria para a descarga dos caminhões da coleta. Os resíduos podem ficar dispostos nestes locais por até vinte e quatro horas e devem ser carregados em caminhões "carreta" (com maior capacidade volumétrica) para serem transportados até o respectivo aterro sanitário. A unidade de transbordo serve para otimizar e reduzir os custos de transporte, e deve possuir licenciamento ambiental para a sua construção, implementação e operação.

2.2 Descoberta de Conhecimento em Bases de Dados (KDD)

A capacidade de detectar padrões válidos, inovadores, vantajosos e coerentes constitui as características do processo de Descoberta de Conhecimento em Bases de Dados (KDD). Nesse estudo considera-se a abordagem apresentada por Fayyad, Piatetsky-Shapiro e Smyth (1996), em que KDD é abordado como um processo composto de uma sequência de etapas, como apresentado na Figura 2. A etapas são: Seleção - etapa onde se define o propósito do projeto para realizar adequadamente a coleta de dados; Pré-processamento - identifica e trata as inconsistências (como: registros incompletos, valores errados e dados incoerentes); Transformação - utiliza técnicas para otimizar o desempenho do modelo na etapa de MD (como: generalização de atributos, discretização de variáveis e normalização); Mineração de Dados - consiste na aplicação de algoritmos de descoberta e análise de dados que constroem modelos sobre os dados; e Análise/Assimilação - executa a validação do modelo por meio da aplicação

de medidas de desempenho e qualidade (como: precisão, erro e confiança) (Fayyad, Piatetsky-Shapiro & Smyth, 1996; Leary *et al.*, 2020; Kvasničková-Stanislavská *et al.*, 2020). Nas seções a seguir estão detalhadas todas as etapas do desenvolvimento desta pesquisa.

Figura 2. Etapas que compõe o processo de KDD



Fonte: Adaptado de Fayyad, Piatetsky-Shapiro e Smyth (1996).

2.2.1 Seleção

A base de dados foi construída com informações dos municípios de Ijuí, Itaqui e São Gabriel, do período de junho de 2020 até outubro de 2020. Por meio da empresa responsável pela coleta de resíduos sólidos urbanos desses municípios, foram obtidos relatórios do sistema de gestão de frotas que monitora todos os caminhões utilizados para a execução do serviço, em dois formatos diferentes apresentados na Figura 3 ((a) - *xlsx* e (b) - *pdf*). Destes relatórios foram extraídos dados como: *Placa dos veículos*; *Estado* (condição física momentânea dos veículos); *Data de Início e Hora de Início* do funcionamento dos veículos, *Data e Hora de Fim*; *Nome do Motorista*; *Ponto de Saída* (localização do caminhão no momento da coleta) e *Peso Líquido* dos resíduos coletados a partir das pesagens dos caminhões.

Figura 3. Exemplo de (a) Relatório I (b) Relatório II do sistema de gestão de frotas.

(a)

Movimentações do veículo IXS-5454-43-CO de 01/08/2020 00:00:00 até 31/08/2020 23:59:59

Estado	Data Início	Hora Início	Data Fim	Hora Fim	Tempo	Dist. Perc (m)	Vel. Máx.	Ponto Inicial	Ponto Final
Parado Ligado	01/08/2020	00:01:36	01/08/2020	00:02:16	40s	0	0 km/h	Ijuí, RS	Ijuí, RS
Andando	01/08/2020	00:02:16	01/08/2020	00:05:09	2m-53s	682	70 km/h	Ijuí, RS	Ijuí, RS
Parado Desligado	01/08/2020	00:05:09	01/08/2020	00:05:15	6s	0	0 km/h	Ijuí, RS	Ijuí, RS
Parado Ligado	01/08/2020	00:05:15	01/08/2020	00:06:49	1m-34s	0	0 km/h	Ijuí, RS	Ijuí, RS
Parado Desligado	01/08/2020	00:06:49	01/08/2020	06:22:15	6h-15m-26s	0	0 km/h	Ijuí, RS	Ijuí, RS

(b)

Tipo de Serviço: COLETA - IJUI

ID	Destino	Nº Cupom	Data	Hora	Veículo	Motorista	Peso Entrada	Peso Saída	Peso Líquido
74804	TRANSBORDO - IJUI	5407	01/08/2020	09:41:00	IXT-9232		17.960,000	12.740,000	5.220,000
74805	TRANSBORDO - IJUI	5408	01/08/2020	09:42:00	IWL-4860		16.620,000	11.080,000	5.540,000
74806	TRANSBORDO - IJUI	5410	01/08/2020	10:58:00	IXS-5454		20.640,000	12.540,000	8.100,000
74807	TRANSBORDO - IJUI	5411	01/08/2020	12:56:00	IXS-7092		20.700,000	12.480,000	8.220,000
74808	TRANSBORDO - IJUI	5412	01/08/2020	13:15:00	IXT-9232		16.560,000	12.500,000	4.060,000

Fonte: empresa responsável pela coleta de resíduos sólidos urbanos dos municípios.

Além disso, foram utilizados dados do último censo demográfico realizado no Brasil pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e do Instituto Nacional de Meteorologia (INMET). Do censo foram coletados: Número de habitantes, Escolaridade, Área total do município, Produto Interno Bruto (PIB) per capita e Índice de Desenvolvimento

Humano (IDHM); e do INMET: Temperatura Média, Temperatura Máxima e Temperatura Mínima. Chhay *et al.* (2018) considera que identificar os fatores que influenciam a geração de RSU é um dos problemas mais importantes e desafiadores na previsão de RSU, por essa razão as variáveis escolhidas foram definidas a partir do estudo da literatura, permitindo identificar os fatores que impactam na geração de resíduos sólidos.

2.2.2 Pré-processamento

Coube nesta etapa trabalhar a compreensão dos dados buscando esclarecer a sua representatividade no processo e a sua contribuição para a descoberta de conhecimento. Para isso, durante o pré-processamento dos dados de coleta de RSU, por meio dos relatórios do sistema de gestão de frotas, identificou-se que um deles foi desenvolvido com registros da posição do caminhão de coleta a cada 30 segundos e o outro com as paradas na unidade de transbordo para descarga temporária dos resíduos. Destes relatórios, é importante esclarecer a interpretação dos dados, dentre eles, o item Estado, que representa o status do caminhão ao longo do percurso e que continha valores que representavam os três estágios que o caminhão poderia se encontrar: Parado Desligado, Parado Ligado e Andando. Outro dado importante é o Ponto Saída, o qual gera um *link* de internet vinculado a um mapa que indica a localização exata do caminhão no momento da coleta.

Na etapa de pré-processamento também foram aplicados filtros com o intuito de preparar a base de dados. Desta forma, foram mantidos apenas os registros que continham o status *Parado Ligado* para o item *Estado*, visto que este status é a identificação de paradas para o recolhimento dos resíduos. Também foram removidos outros itens disponíveis nos relatórios considerados irrelevantes, como: *Tempo*, *Distância Percorrida* (m), *Velocidade Máxima*, *Ponto Inicial*, *ID*, *Destino*, *Número Cupom*, *Peso Entrada* e *Peso Saída*. As atividades realizadas nesta etapa foram realizadas com apoio de recursos do software Excel.

2.2.3 Transformação

Nesta etapa, para otimizar o desempenho do modelo para a etapa de MD, os itens *Data Início*, *Hora Início*, *Data Fim*, *Hora Fim* foram agrupadas em *Data Hora Início* e *Data Hora Fim*, com o objetivo de transformá-los em um item único, o que permitiu o cálculo correto do tempo entre a *Data Hora Fim* e *Data Hora Início*, sem inconsistência para os registros em que a coleta ocorreu na madrugada durante a passagem de um dia para o outro. A partir dessa subtração foi adicionado o item *Duração*, representando o tempo em segundos que o caminhão fica parado para coleta dos resíduos. Relacionando logicamente o item *Estado* com o item *Duração*, também foram removidos da *Duração* os valores maiores ou iguais 10 minutos com o status *Parado Ligado*, em justificativa do caminhão estar sendo abastecido de combustível ou estar recebendo manutenção corretiva.

O item *Ponto Final* foi convertido inicialmente para o tipo *Texto* com o objetivo de extrair do *link* as coordenadas correspondentes a cada ponto de coleta. Com a extração dos valores de latitude e longitude, foram criados dois novos itens na base de dados (*Latitude* e *Longitude*).

O cálculo dos pesos dos RSU coletados em cada ponto foi obtido a partir dos itens *Placa*, *Data Hora Início*, *Data Hora Fim*, *Duração* e *Peso Líquido*. Para isso, foi realizada a divisão proporcional relacionando o *Peso Líquido* de resíduos aferido ao término de cada trajeto e o tempo de parada do veículo em cada ponto de coleta do respectivo trajeto. Para assegurar a fidelidade da distribuição dos pesos ao longo dos percursos de coleta, foram identificados e subtraídos os tempos das paradas dos veículos realizadas para a descarga dos resíduos nas unidades de transbordo.

2.2.4 Mineração de Dados

A mineração de dados tem a finalidade de descobrir padrões em dados de forma automática ou semiautomática (Kantardzic, 2011; Witten, Frank & Hall, 2011; Kolling et al., 2021). Para isso e por meio da revisão da literatura, foram selecionados os algoritmos utilizados para descoberta de conhecimento em processos do gerenciamento de RSU: Regressão Linear (RL); *Gradient Boosting (GB)* e *Random Forest (RF)* (Azadi & Karimi-Jashni, 2016; Ceylan, 2020; Chhay et al., 2018; Golbaz, Nabizadeh & Sajadi, 2019; Hartnett, Ayyasamy & Balachandran, 2019). Para análise foram utilizadas as seguintes variáveis explicativas: Placa, Data Hora Início, Data Hora Fim, Cidade, Latitude, Longitude, Área Territorial (km²), População estimada (hab), Densidade Demográfica (hab/km²), Escolarização 6 a 14 anos, IDHM, PIB per Capita, Temperatura Média (°C), Temperatura Mínima (°C), Temperatura Máxima (°C) e Chuva (mm), sendo o Volume Coletado (Kg) por parada a variável dependente.

As variáveis foram normalmente distribuídas com média zero e variância constante, uma vez que para as equações de regressão a colinearidade pode implicar em resultados inconvenientes devido ao fato de que altas correlações entre variáveis preditoras podem dificultar a análise correta (Sousa et al., 2007; Azadi & Karimi-Jashni, 2016). Também foi aplicada a técnica de Análise de Componentes Principais (PCA) para remover a dependência das variáveis explicativas (Azadi & Karimi-Jashni, 2016). Essa técnica trata-se de um método estatístico multivariado amplamente empregado para criar novas variáveis que não são correlacionadas. Essas variáveis independentes são combinações lineares das variáveis originais e contêm as informações das variáveis originais com perdas mínimas (Noori et al., 2010; Azadi & Karimi-Jashni, 2016). Além disso, para evitar possível viés na seleção dos indivíduos do conjunto de dados, a etapa de treinamento foi realizada por meio do método de validação cruzada, dividindo a amostra em 10 conjuntos de dados para treinamento e teste.

Para a MD foi utilizado o *software* Orange, que é um pacote de aprendizado de máquina para análise de dados por meio de *scripts* Python e programação visual (Demšar et al., 2013). Além da análise do modelo de regressão, também foi realizada uma análise descritiva do conjunto de dados, a fim de caracterizar o contexto e os resultados obtidos (Colvero et al, 2019). Os dados coletados para a amostra também foram analisados extensivamente através do *software* Microsoft Power BI.

2.2.4.1 Regressão Linear Múltipla

A Regressão Linear Múltipla (RLM) envolve duas ou mais variáveis explicativas ou independentes que impactam uma única variável dependente, sendo modeladas ajustando uma equação linear aos dados de aprendizagem (Azadi & Karimi-Jashni, 2016). É usual empregar a RLM para previsão de geração de RSU, uma vez que esse contexto também é afetado por vários fatores, permitindo que o modelo RLM estabeleça a relação entre as variáveis independentes e a variável dependente (Hartnett, Ayyasamy & Balachandran, 2019).

2.2.4.2 Random Forest

O Random Forest (RF) é um método de aprendizagem por conjunto que resulta em um número especificado de árvores de decisão imparciais que agrega todas as decisões e faz a previsão final. Estas árvores são obtidas pelo método *bootstrap* e a maioria dos seus algoritmos retém em torno de 30% dos dados de treinamento durante o processo. Os pontos de dados salvos são utilizados para avaliar a precisão do modelo e são denominados de *out-of-bag* (OOB). O algoritmo cria um número pré-especificado de amostras *bootstrap* (*number of trees to Grow - NTREE*), onde quanto mais elevados os valores, maior é a precisão do modelo. Posteriormente, o algoritmo apresenta um número pré-especificado de parâmetros de entrada (*number of*

variables randomly sampled as candidates at each split - MTRY) e gera uma árvore de decisão para cada amostra com base nestes parâmetros (Hartnett, Ayyasamy & Balachandran, 2019).

2.2.4.3 Gradient Boosting

Este algoritmo cria uma árvore de decisão durante o processo de aprendizado, desenvolvendo o processo sequencialmente, característica pela qual se diferencia do RF. A primeira árvore é montada aleatoriamente e durante a execução vai ajustando os erros do modelo a partir da análise dos dados. A previsão final da classe é estimada como a soma das previsões das árvores individuais. É indispensável a definição de três parâmetros de ajuste: (a) o número de árvores, (b) o número máximo de divisões em cada árvore e (c) a taxa de aprendizagem do gradiente da função de perda. Para isso, foi fixado o número máximo de divisões em cada árvore e os hiper-parâmetros da taxa de aprendizagem para os valores padrão, bem como ajustado o número de árvores (Hartnett, Ayyasamy & Balachandran, 2019).

2.2.5 Análise/Assimilação

Para a regressão, a validação do modelo para previsão de RSU é aplicada por meio das medidas de desempenho. Para este trabalho foram adotados: erro médio absoluto (*Mean absolute error - MAE*), erro quadrático médio (*Root Mean Squared Error - RMSE*) e coeficiente de determinação (R^2). O alto valor de R^2 e os baixos valores de MAE e RMSE significam que o regressor desenvolvido tem melhor desempenho.

A Equação (1) apresenta como foi calculado o MAE, o qual resulta no valor médio entre valores observados (reais) e as previsões (hipóteses).

$$MAE = \frac{1}{n} \sum_{i=1}^n |dj - pj| \quad (1)$$

A Equação (2) foi usada para o RMSE, que é a medida que calcula "a raiz quadrática média" dos erros entre valores observados (reais) e previsões (hipóteses).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (dj - pj)^2}{n}} \quad (2)$$

A Equação (3), determina o Coeficiente de determinação (R^2), o qual é uma medida estatística de quão próximos os dados estão da linha de regressão ajustada. Ele também é conhecido como o coeficiente de determinação ou o coeficiente de determinação múltipla para a regressão múltipla.

$$R^2 = \left[\frac{\sum_{i=1}^n \frac{(dj - \bar{d})(pj - \bar{p})}{\sqrt{\sum_{i=1}^n (dj - \bar{d})^2 \times \sum_{i=1}^n (pj - \bar{p})^2}}}{\sum_{i=1}^n} \right] \quad (3)$$

Onde dj é a saída real para o j -ésimo valor, pj é a saída prevista para o j -ésimo valor, d e p são a média de toda a produção real e prevista e n representa o número de saídas variáveis.

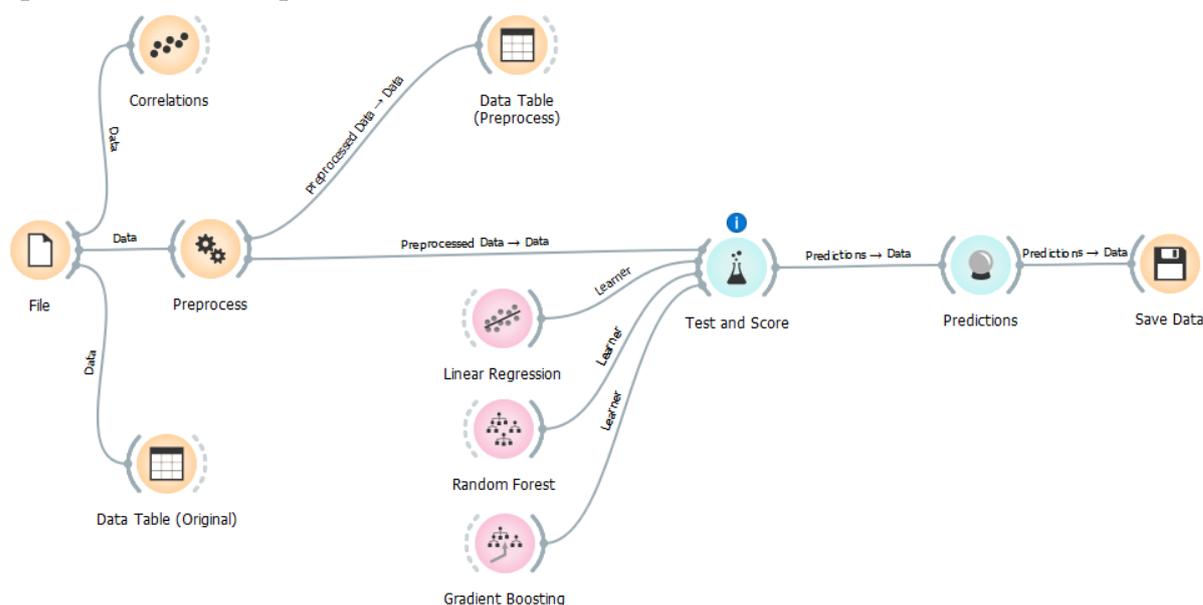
3 Resultados e discussões

Em 2019, o Brasil gerou 79,06 milhões de toneladas de resíduos, o que representa um aumento de 18,65% em comparação com 2010 (ABRELPE, 2021). Os resíduos gerados pela sociedade exigem a atenção dos governos municipais por meio da implementação de estratégias de curto, médio e longo prazo, uma vez que a parcela de recursos públicos ou privados destinados ao gerenciamento de resíduos é insuficiente para fornecer uma infraestrutura que minimize as deficiências. Portanto, soluções que reduzam custos e melhorem a qualidade da coleta de resíduos, gestão do transporte e disposição final são essenciais para melhorar o ciclo. Neste contexto, o uso de técnicas computacionais se mostra como uma ferramenta eficaz para otimizar o gerenciamento da coleta de resíduos sólidos, tornando o processo mais produtivo e sustentável, com capacidade para reduzir distâncias excessivas na rota, permitir o máximo transporte de resíduos, diminuir o desperdício, aumentar a eficiência energética e minimizar as emissões geradas (Dias et al., 2021). Neste sentido, há uma crescente demonstração de preocupação em todos os níveis da sociedade para desenvolver estratégias específicas de gestão municipal que possam proporcionar melhores resultados contemplando os aspectos ambientais, sociais e econômicos (Kupusamy et al., 2019; Sott et al., 2020; Furstenau et al., 2020).

O processo de coleta e transporte de resíduos sólidos é um aspecto crucial do gerenciamento de resíduos, pois impacta diretamente a eficiência e a eficácia da disposição dos resíduos. A coleta e transporte adequados dos resíduos garantem que sejam manuseados e descartados de maneira que minimize os riscos ambientais e de saúde. Práticas inadequadas de coleta e transporte podem levar à contaminação do solo, água e ar, bem como à propagação de doenças. Além disso, sistemas de coleta e transporte ineficientes podem resultar em aumento de custos e redução da satisfação pública com os serviços de gerenciamento de resíduos. Portanto, é essencial analisar o processo de coleta e transporte de resíduos sólidos para identificar áreas de melhoria e otimizar as práticas de gestão de resíduos.

A fim de implementar o modelo de Regressão Linear, *Gradient Boosting* e *Random Forest* para estimar e prever a coleta de RSU por ponto, nos municípios de Ijuí, Itaqui e São Gabriel, foi utilizado o *software* Orange. Na Figura 4, por meio da programação visual da ferramenta podem ser observadas as etapas metodológicas.

Figura 4. Modelo no Orange



Fonte: Elaborado pelos autores (2024).

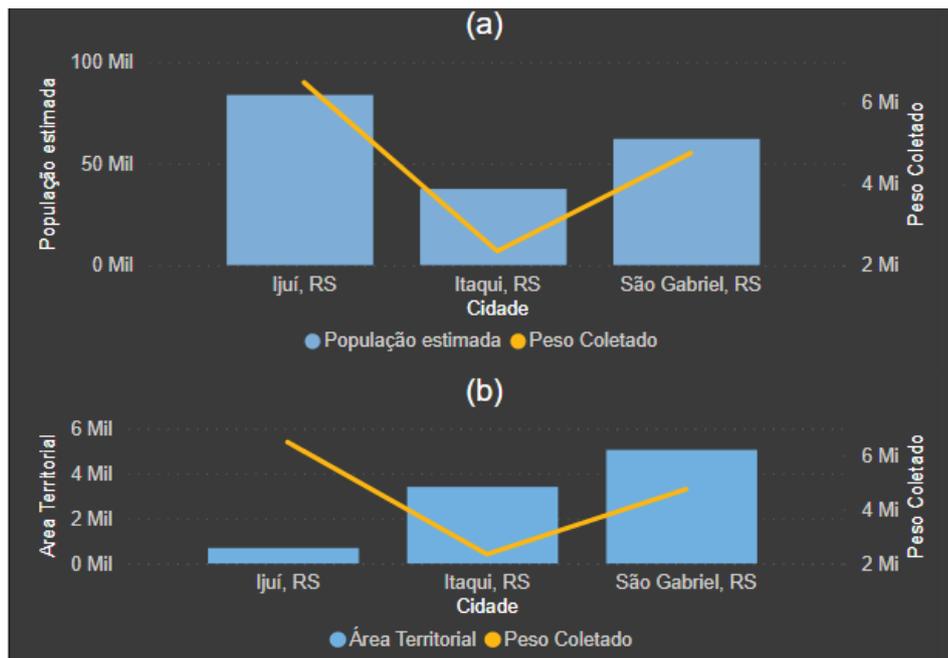
A partir do tratamento dos dados foi possível observar que a análise contém informações das três (3) cidades, com o total de dez (10) veículos de coleta (5 em Ijuí, 2 em Itaqui e 3 em São Gabriel), e vinte e quatro (24) motoristas para atender a todos os turnos de trabalho. A Figura 5 exemplifica por meio da visão da Rua Ernesto Alves com a Rua José Bonifácio, no município de Ijuí, onde foi realizada a análise de todos os pontos em que os caminhões da coleta executaram paradas no período da amostragem de junho de 2020 até outubro de 2020.

Figura 5. Visão do município de Ijuí mais o recorte da Rua Ernesto Alves com a Rua José Bonifácio.



Por meio dos dados analisados, descobriu-se que a média do volume coletado por segundo é equivalente a 2,13kg. A relação entre a população estimada e o volume recolhido representada na Figura 6(a) confirma que o número de habitantes é um dos fatores que influencia os índices de coleta de resíduos. Já a relação inversa entre a área do município e peso coletado é representada na Figura 6(b) e demonstra que nesse conjunto de dados o tamanho da cidade não impactou a coleta.

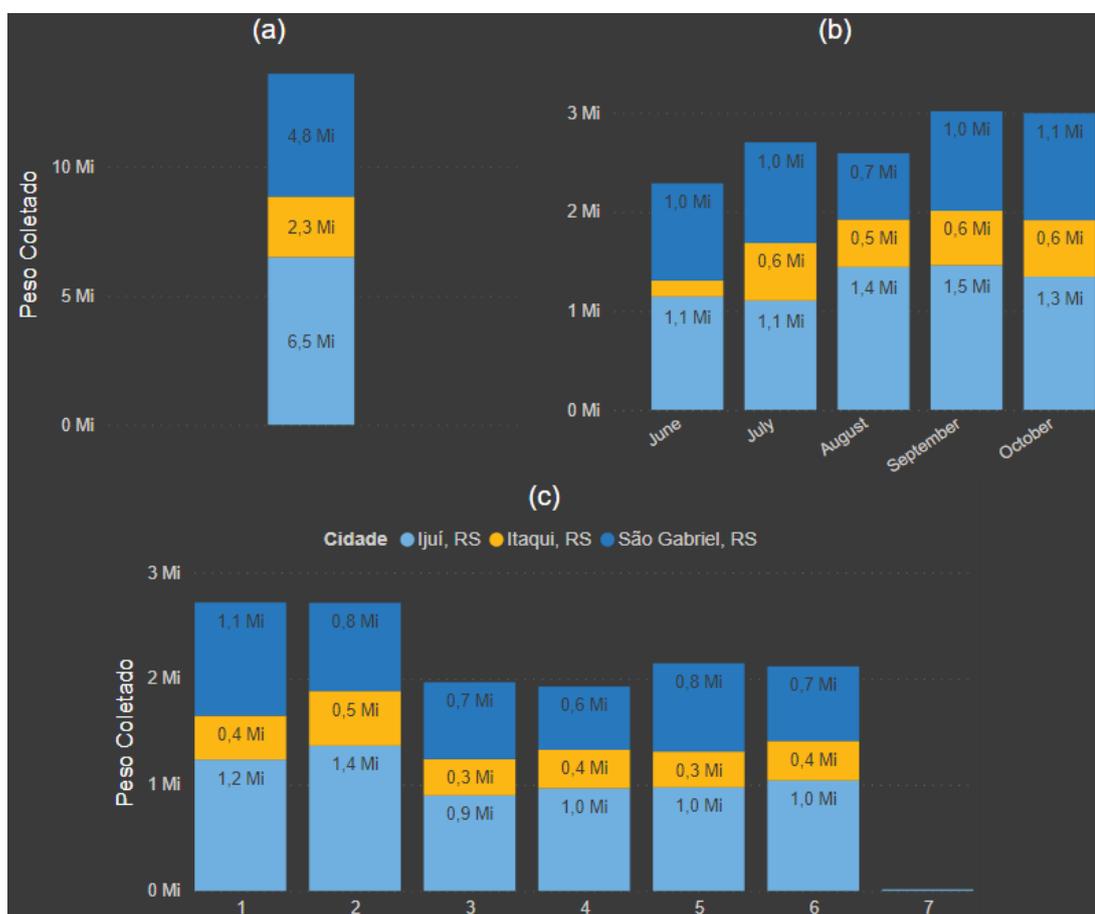
Figura 6. (a) População estimada x Volume coletado (b) Área do município x Volume coletado



Fonte: Elaborado pelos autores (2024).

A Figura 7 apresenta o panorama da coleta dos RSU dos municípios de Ijuí, Itaqui e São Gabriel dos 5 meses de amostragem (junho a outubro de 2020). A Figura 7a demonstra que foram coletados os totais de 6,5 mil ton, 2,3 mil ton e 4,8 mil ton respectivamente em Ijuí, Itaqui e São Gabriel. A partir destes valores, obteve-se que a coleta per capita desses municípios, no período de amostragem foi de 0,51 kg/hab/dia em Ijuí; 0,40 kg/hab/dia em Itaqui e 0,50 kg/hab/dia em São Gabriel, ficando abaixo da coleta per capita média (0,72 kg/hab/dia) aferida para os municípios do Estado do Rio Grande do Sul pela ABRELPE (2021). A Figura 7b corrobora a representatividade da Figura 6a, de que o número de habitantes é proporcional ao índice de geração e coleta dos RSU, portanto em Ijuí é coletado mais resíduos do que em São Gabriel, e em São Gabriel são coletados mais resíduos do que em Itaqui.

Figura 7. (a) Peso Total Coletado por cidade durante os cinco meses (b) Peso Coletado x Meses (c) Quantidade de resíduos coletados (ton/dia da semana)



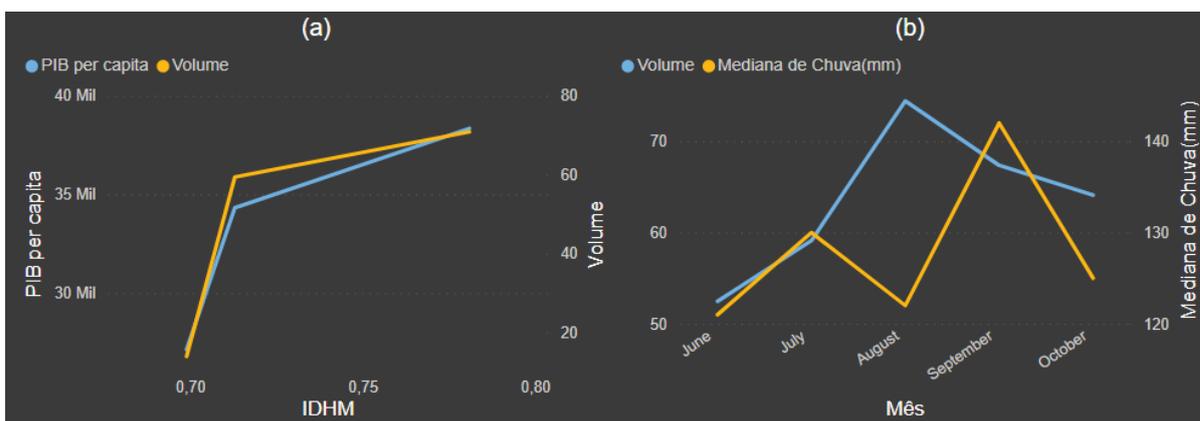
Fonte: Elaborado pelos autores (2024).

A Figura 7c identifica para os três municípios o dia de semana em que é coletada a maior quantidade de RSU. Os dias da semana são representados como segunda-feira: 1, terça-feira: 2, quarta-feira: 3, quinta-feira: 4, sexta-feira: 5, sábado: 6 e domingo: 7. O domingo só foi identificado para o município de Ijuí, podendo ter ocorrido coletas para concluir algum roteiro da semana devido a problemas operacionais, justificando a pequena quantidade coletada. Em Ijuí e Itaqui a terça-feira é o dia com maior quantidade coletada, totalizando no período 1,37 mil toneladas e 0,51 mil toneladas, respectivamente. Tanto para Ijuí, quanto para Itaqui a tendência de quantidade de resíduos coletada é a mesma ao longo da semana, seguindo com segunda-feira e sábado. Para São Gabriel, o dia com maior quantidade de resíduos coletados é

segunda-feira com 1,07 mil toneladas, tendo terça-feira e sexta-feira na sequência com a mesma quantidade coletada de 0,83 mil toneladas. Essa variação ocorre devido a setorização dos bairros que é feita em cada município, a qual separa a periodicidade em alternada "par" (segundas, quartas e sextas-feiras) e "ímpar" (terças, quintas-feiras e sábados). Assim os caminhões percorrem os mesmos trajetos e pontos de coleta de acordo com a alternada correspondente.

A Figura 8(a) mostra a relação de proporcionalidade entre as variáveis explicativas independentes de PIB per capita e IDHM com a variável dependente de Volume Coletado de RSU levando em consideração o somatório dos três municípios. A Figura 8(b) ressalta a não existência de relação direta entre a variável explicativa independente Mediana de Chuva e a variável dependente de Volume Coletado. Esta análise buscou visualizar a interferência que a chuva exerce devido aos municípios disporem seus resíduos nas frentes das residências (lixeiros comuns ou na calçada), visto que quando mal acondicionados ou até mesmo quando colocados nos caminhões compactadores, os RSU absorvem a água precipitada, interferindo na pesagem realizada antes de serem descarregados na unidade de transbordo.

Figura 8. Relação das variáveis independentes e a variável dependente: (a) PIB per capita e IDHM x Volume coletado (b) Mediana de Chuva x Volume coletado ao longo dos meses



Fonte: Elaborado pelos autores (2024).

Para a descoberta de conhecimento em processos do gerenciamento dos RSU, a Tabela 1 evidencia a comparação dos algoritmos aplicados de Regressão Linear (RL), *Gradient Boosting* (GB) e *Random Forest* (RF). Nesta comparação ficou evidenciado que o modelo de Regressão Linear apresentou desempenho de MAE: 50,942, RMSE: 134,891 e R²: 0,152; o modelo de *Gradient Boosting* teve como desempenho de MAE: 25,244; RMSE: 87,667 e R²: 0,642; já o modelo de *Random Forest* teve como desempenho de MAE: 19,740, RMSE: 88,826 e R²: 0,632.

Tabela 1. Comparação entre modelos

Modelo	MAE	RMSE	R2
Regressão Linear	50,942	134,891	0,152
<i>Gradient Boosting</i>	25,244	87,667	0,642
<i>Random Forest</i>	19,740	88,826	0,632

Considerando que o regressor desenvolvido com o melhor desempenho deve apresentar valores baixos de MAE e RMSE e valor alto de R^2 e comparando os modelos de RL, GB e RF, o menor valor de MAE foi obtido pelo RF, o menor valor de RMSE foi alcançado pelo GB e o maior valor de R^2 foi aferido também pelo GB. Visto isso, conclui-se que o modelo de algoritmo *Gradient Boosting* apresentou os melhores resultados para processos de gerenciamento dos RSU.

Os valores obtidos para MAE, RMSE e R^2 não são valores ótimos, mas expressam grande potencial para desenvolvimento, cabendo ressaltar o forte impacto da escassez de dados e a complexidade da coleta e tratamento na execução do processo de descoberta de conhecimento. A escassez de dados percebida neste estudo se dá devido à restrição de tempo de armazenamento de dados do software de gestão de frotas impossibilitando utilizar um período maior para treinamento e teste do modelo (limitado a 6 meses de dados sendo que na data da coleta apenas 5 meses puderam ser coletados). Sugere-se que a coleta seja realizada no período de um ano para avaliar e comparar o comportamento do modelo com maior volume de dados. Além disso, a complexidade do tratamento dos dados é considerada a tarefa mais importante do processo e onde o pesquisador dedica o período mais longo de tempo. Nesse estudo o período de tratamento de dados, estimado em 80% por alguns autores, foi ultrapassado e chegou a 85.38% (equivalente a 13.5 meses) do tempo total (16 meses) dedicado ao processo de descoberta de conhecimento.

4 Conclusão

Por meio do processo de KDD e das técnicas de mineração de dados, este trabalho avaliou a geração de resíduos sólidos urbanos em três municípios do estado do Rio Grande do Sul. Para isso, foi utilizado o histórico de dados acerca do peso de descarga dos caminhões de coleta, das rotas de recolhimento, dados sociodemográficos e de clima. Após a coleta, os dados foram tratados e posteriormente analisados para prever a quantidade de resíduos sólidos gerados por ponto de coleta. As métricas de avaliação do modelo de predição foram utilizadas para avaliar a qualidade da base de dados e do modelo, cujos resultados expressam grande potencial para desenvolvimento e merecem investimento de estudos mais extensos para apresentar resultados ainda mais significativos.

Este estudo contribui para a literatura acadêmica ao reforçar a constatação de que o PIB per capita e o IDHM apresentam uma relação direta na geração de RSU e, conseqüentemente, nos níveis de coleta. Além disso, apresenta uma visão divergente da literatura ao afirmar que o nível de precipitação não afeta diretamente o volume de resíduos coletados nos três municípios analisados. Este estudo contribui para a prática ao levantar questões para gestores municipais sobre o gerenciamento de resíduos e as variáveis que impactam sua coleta. Estudos futuros podem adicionar mais variáveis à análise e utilizar um volume maior de dados sobre coleta de resíduos. Adicionalmente, uma análise qualitativa pode ser conduzida para compreender os aspectos da gestão municipal que impactam o processo.

Nesse sentido, esta pesquisa gera contribuição em dois sentidos: primeiro, gera informações que podem auxiliar na tomada de decisão organizacional e melhorar o serviço de coleta oferecido à comunidade local em termos sociais e ambientais. Segundo, este estudo colabora com a literatura acadêmica ao reforçar o potencial da utilização da mineração de dados para o gerenciamento de resíduos sólidos urbanos, reforçando o potencial de tecnologias emergentes para solucionar problemas regionais. Neste trabalho, um grande volume de dados foi explorado para analisar e comparar informações de três cidades. Outras técnicas podem ser incorporadas no futuro para análises mais aprofundadas.

Além disso, como continuação deste estudo, pode-se utilizar os resultados da análise preditiva como parâmetro de entrada para desenvolver uma ferramenta que programe a rota

otimizada para o veículo de coleta e transporte dos resíduos municipais, e a partir da análise descritiva, produzir informações relevantes para o poder público municipal embasar a tomada de decisão e elaborar um planejamento eficaz para o gerenciamento de resíduos sólidos urbanos, além de contribuir para o crescimento do engajamento da população, aprimorar a educação ambiental nos municípios, reduzir custos de coleta e destinação dos RSU. Assim, contemplando o tripé da sustentabilidade (aspectos sociais, econômicos e ambientais) e adicionando uma importante hélice da sociedade moderna, que é a tecnologia.

Ainda pode-se usar a mineração de dados em outros processos do gerenciamento de resíduos sólidos que não compõem diretamente a coleta e roteamento de veículos, como modelos que auxiliem nos processos de logística reversa e no destino correto dos resíduos, a fim de reduzir os efeitos da destinação imprópria na vida de milhares de pessoas e animais do planeta. Além disso, a integração com outras tecnologias, como robôs autônomos para coleta e transporte e sensores para classificação, podem transformar a forma de coletar e tratar resíduos sólidos urbanos.

Referências

- Abbasi, M., Rastgoo, M. N., & Nakisa, B. (2019). Monthly and seasonal modeling of municipal waste generation using radial basis function neural network. *Environmental Progress & Sustainable Energy*, 38(3), e13033.
- ABRELPE (2021). Panorama dos Resíduos Sólidos no Brasil 2020. Associação Brasileira de Empresas de Limpeza Pública e Resíduos Especiais. São Paulo.
- Ahmad, S., & Kim, D. H. (2020). Quantum GIS based descriptive and predictive data analysis for effective planning of waste management. *IEEE Access*, 8, 46193-46205.
- Ahmed, M. S., Arif, M. F., & Hossain, M. M. (2020). Prediction of solid waste generation and finding the sustainable pathways in the city of Dhaka. *Management of Environmental Quality: An International Journal*.
- Azadi, S., & Karimi-Jashni, A. (2016). Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of Fars province, Iran. *Waste management*, 48, 14-23.
- IBGE (2021). Conheça Cidades e Estados do Brasil. Instituto Brasileiro de Geografia e Estatística. Disponível em: <https://cidades.ibge.gov.br>. Acessado em: março, 2021.
- Brasil (2010). Política Nacional dos Resíduos Sólidos, Lei Nº 12.305, de 2 de agosto de 2010. Diário Oficial da República, 1º série, nº 116, Brasília.
- Buenrostro-Delgado, O., Ortega-Rodriguez, J. M., Clemitshaw, K. C., González-Razo, C., & Hernández-Paniagua, I. Y. (2015). Use of genetic algorithms to improve the solid waste collection service in an urban area. *Waste management*, 41, 20-27.
- Ceylan, Z., Bulkan, S., & Elevli, S. (2020). Prediction of medical waste generation using SVR, GM (1,1) and ARIMA models: a case study for megacity Istanbul. *Journal of Environmental Health Science and Engineering*, 11. doi:10.1007/s40201-020-00495-8
- Ceylan, Z. (2020). Estimation of municipal waste generation of Turkey using socio-economic indicators by Bayesian optimization tuned Gaussian process regression. *Waste Management & Research*, 38(8), 840-850.
- Chhay, L., Reyad, M. A. H., Suy, R., Islam, M. R., & Mian, M. M. (2018). Municipal solid waste generation in China: influencing factor analysis and multi-model forecasting. *Journal of Material Cycles and Waste Management*, 20(3), 1761-1770.
- Colvero, D. A., Feitosa, A. K., Ramalho, J. C., Gomes, A. P. D., Tarelho, L. A. D., & de Matos, M. A. A. (2019). Per capita municipal solid waste generation and its relationship with socioeconomic and demographic factors in Goiás State, Brazil. *Revista Tecnologia E Sociedade*, 15(36), 254-+. doi:10.3895/rts.v15n36.8757

- Coskuner, G., Jassim, M. S., Zontul, M., & Karateke, S. (2020). Application of artificial intelligence neural network modeling to predict the generation of domestic, commercial and construction wastes. *Waste Management & Research*, 9. doi:10.1177/0734242x20935181
- Dai, F., Nie, G. H., & Chen, Y. (2020). The municipal solid waste generation distribution prediction system based on FIG-GA-SVR model. *Journal of Material Cycles and Waste Management*, 22(5), 1352-1369. doi:10.1007/s10163-020-01022-5
- Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., ... & Zupan, B. (2013a). Orange: data mining toolbox in Python. *the Journal of machine Learning research*, 14(1), 2349-2353.
- Demšar, J., & Zupan, B. (2013b). Orange: Data mining fruitful and fun-a historical perspective. *Informatica*, 37(1).
- Dias, J. L., Sott, M. K., Ferrão, C. C., Furtado, J. C., & Moraes, J. A. R. (2021). Data mining and knowledge discovery in databases for urban solid waste management: A scientific literature review. *Waste Management & Research*, 39(11), 1331-1340.
- Di Maria, F., Micale, C., & Morettini, E. (2016). Impact of the pre-collection phase at different intensities of source segregation of bio-waste: An Italian case study. *Waste management*, 53, 12-21.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. Paper presented at the KDD.
- Ferrão, C. C.; Moraes, J. A. R. Analysis of environmental risks and accidents at work in urban solid waste collection services. *Gestão & Produção*, v. 28, 2021. ISSN 0104-530X.
- Furstenau, L. B., Sott, M. K., Kipper, L. M., Machado, E. L., Lopez-Robles, J. R., Dohan, M. S., ... & Imran, M. A. (2020). Link between sustainability and industry 4.0: trends, challenges and new perspectives. *Ieee Access*, 8, 140079-140096.
- Guerrini, A., Carvalho, P., Romano, G., Marques, R. C., & Leardini, C. (2017). Assessing efficiency drivers in municipal solid waste collection services through a non-parametric method. *Journal of cleaner production*, 147, 431-441.
- Golbaz, S., Nabizadeh, R., & Sajadi, H. S. (2019). Comparative study of predicting hospital solid waste generation using multiple linear regression and artificial intelligence. *Journal of Environmental Health Science and Engineering*, 17(1), 41-51.
- Gouveia, N. (2012). Resíduos sólidos urbanos: impactos socioambientais e perspectiva de manejo sustentável com inclusão social. *Ciência & saúde coletiva*, 17, 1503-1510.
- Hannan, M. A., Akhtar, M., Begum, R. A., Basri, H., Hussain, A., & Scavino, E. (2018). Capacitated vehicle-routing problem model for scheduled solid waste collection and route optimization using PSO algorithm. *Waste management*, 71, 31-41.
- Hartnett, T. Q., Ayyasamy, M. V., & Balachandran, P. V. (2019). Prediction of new iodine-containing apatites using machine learning and density functional theory. *MRS Communications*, 9(3), 882-890.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.
- Kinobe, J. R., Bosona, T., Gebresenbet, G., Niwagaba, C. B., & Vinnerås, B. (2015). Optimization of waste collection and disposal in Kampala city. *Habitat International*, 49, 126-137
- Kolling, M. L., Furstenau, L. B., Sott, M. K., Rabaioli, B., Ulmi, P. H., Bragazzi, N. L., & Tedesco, L. P. C. (2021). Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development. *International journal of environmental research and public health*, 18(6), 3099.
- Kontokosta, C. E., Hong, B., Johnson, N. E., & Starobin, D. (2018). Using machine learning and small area estimation to predict building-level municipal solid waste generation in

- cities. *Computers Environment and Urban Systems*, 70, 151-162. doi:10.1016/j.compenvurbsys.2018.03.004
- Kumar, A., & Samadder, S. R. (2017). An empirical model for prediction of household solid waste generation rate—A case study of Dhanbad, India. *Waste Management*, 68, 3-15.
- Kupusamy, K., Nagapan, S., Abdullah, A., Kaliannan, S., Sohu, S., Subramaniam, S., & Maniam, H. (2019). Construction Waste Estimation Analysis in Residential Projects of Malaysia. *Engineering Technology & Applied Science Research*, 9(5), 4842-4845.
- Kvasničková Stanislavská, L., Pilař, L., MargarISOVÁ, K., & Kvasnička, R. (2020). Corporate Social Responsibility and social media: Comparison between developing and developed countries. *Sustainability*, 12(13), 5255.
- Leary, A., Cook, R., Jones, S., Radford, M., Smith, J., Gough, M., & Punshon, G. (2020). Using knowledge discovery through data mining to gain intelligence from routinely collected incident reporting in an acute English hospital. *International journal of health care quality assurance*.
- Niska, H., & Serkkola, A. (2018). Data analytics approach to create waste generation profiles for waste management and collection. *Waste Management*, 77, 477-485.
- Noori, R., Karbassi, A., & Sabahi, M. S. (2010). Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management*, 91(3), 767-771.
- Pires, A., Sargedas, J., Miguel, M., Pina, J., & Martinho, G. (2017). A case study of packaging waste collection systems in Portugal—Part II: Environmental and economic analysis. *Waste management*, 61, 108-116.
- Ripa, M., Fiorentino, G., Vacca, V., & Ulgiati, S. (2017). The relevance of site-specific data in Life Cycle Assessment (LCA). The case of the municipal solid waste management in the metropolitan city of Naples (Italy). *Journal of Cleaner Production*, 142, 445-460.
- Sott, M. K., Furstenuau, L. B., Kipper, L. M., Giraldo, F. D., Lopez-Robles, J. R., Cobo, M. J., ... & Imran, M. A. (2020). Precision techniques and agriculture 4.0 technologies to promote sustainability in the coffee sector: state of the art, challenges and future trends. *IEEE Access*, 8, 149854-149867.
- Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, 22(1), 97-103.
- Yay, A. S. E. (2015). Application of life cycle assessment (LCA) for municipal solid waste management: a case study of Sakarya. *Journal of Cleaner Production*, 94, 284-293.
- Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, 31(1), 76-77.