

Análise do consumo de combustível da PM-PE com foco no Controle Interno

Ana Catarina G. F. Pereira ¹

Hitalo Oliveira da Silva ¹

Hugo de Andrade Amorim Neto ¹

João Antônio da Silva Lima ¹

Messias Rafael Batista ¹

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

Resumo

A Polícia Militar de Pernambuco (PM-PE) tem um órgão de auditoria e controle interno - a Comissão Permanente de Auditoria (CPA) - que faz a análise de eficiência e conformidade orçamentária da PM-PE. Porém são gerados muitos dados mensalmente, de todas as unidades da PM-PE, tornando por vezes o trabalho manual muito custoso. Para auxiliar no controle interno, este trabalho busca descrever a implementação de uma ferramenta inicial de previsão de consumo de combustível das viaturas. Foi utilizado um modelo de previsão para dados com médias móveis e sazonalidade, o SARIMA, e foi alcançada uma taxa de acerto de 84.43%.

Palavras-Chave: Previsão; Consumo de Combustível; Controle Interno

1 Introdução

1.1 Contextualização

A Comissão Permanente de Auditoria (CPA) foi criada por meio do decreto estadual nº 24.629/2012, como órgão técnico de assessoria direta ao Comando Geral da Polícia Militar de Pernambuco (PM-PE).

O principal objetivo da CPA é avaliar o desempenho da Gestão Operacional por meio da análise de dados e da criação de indicadores de eficiência, que são baseados nas leis e normas do Tribunal de Contas, do Ministério Público e da Controladoria Geral do Estado de Pernambuco. Atualmente as ações de auditoria são divididas por gestões ou competências de auditoria sendo elas: pessoal, material bélico, financeira, combustível, viaturas, patrimônio e almoxarifado.

1.2 Descrição do Problema

O processo de análise da eficiência e da conformidade orçamentária da corporação é realizado atendendo a um plano anual ou por demanda específica de uma unidade policial.

O reconhecimento de possíveis anomalias e desvios é difícil e complexo, pois a aferição é executada manualmente em um grande volume de dados, com o auxílio de planilhas eletrônicas e baseando-se na subjetividade e experiência técnica dos avaliadores. Na área de controle do consumo e gastos com combustível são 82 unidades de gestão que mensalmente enviam relatórios de consumo e gastos para o setor de gestão de combustível da PM-PE.

Um dos pontos verificados durante a auditoria é o consumo de combustível das viaturas da corporação por quilometragem percorrida. Elas são divididas por batalhões e apresentam especificidades com base no modelo do veículo, motorização, região atendida, data de manutenção, condutor, etc.

Os auditores buscam, também de forma manual e subjetiva, indícios de não conformidade, baseando-se no comportamento histórico do consumo de combustível por batalhão e por viatura.

1.3 Objetivo

Implementar uma ferramenta que realize a previsão dos dados de consumo de combustível dos batalhões da PM-PE. O resultado final esperado é de que a CPA disponha de uma ferramenta de tecnologia da informação que possa auxiliar na previsão do consumo de combustível das viaturas e assim auxiliar no controle interno e planejamento.

1.4 Justificativa

A análise dos dados da Gestão da PM-PE é executada mensalmente; de forma manual e subjetiva; em 82 Unidades de Gestão (Batalhões); numa grande quantidade de dados, incluindo o consumo de combustível; e por um quadro reduzido de auditores da CPA. É comum que o intervalo de tempo entre auditorias para um mesmo batalhão extrapole o prazo de um mês, devido ao volume de dados. Portanto, uma ferramenta de previsão do consumo de combustível auxiliará o trabalho dos auditores da CPA, pois permite verificar inconsistências, diminuindo o tempo e os esforços entre as ações de auditorias.

1.5 Escopo Negativo

A ferramenta proposta não tem como objetivo inferir se existe a não conformidade nos dados do consumo de combustível. A avaliação final continua sendo realizada pelo auditor da CPA.

2 Fundamentação Teórica

2.1 Área de Negócio

O policiamento do estado de Pernambuco é realizado pela Polícia Militar deste mesmo estado - PM-PE, sendo este o órgão responsável pela segurança pública visando assegurar o cumprimento da lei, da manutenção da ordem pública assim como o exercício dos poderes a este constituídos, conjuntamente com os demais órgãos e policiais do Estado.

Como todos os órgãos públicos, a Polícia Militar recebe recursos do Estado para exercer sua função. Buscando sempre a otimização e fiscalização dos recursos públicos, é feita uma auditoria com os dados da polícia, mais especificamente com os dados de combustível consumido pelas viaturas dos batalhões da PM-PE. "Auditoria é uma verificação ou exame feito por um auditor dos documentos de

prestação de contas com o objectivo de o habilitar a expressar uma opinião sobre os referidos documentos de modo a dar aos mesmos a maior credibilidade" (TC, 1999, p. 30).

Desta forma, emerge a necessidade da instituição estar sempre preparada com relatórios gerenciais para tomada de decisão, tendo em vista que cada veículo envolve uma série de atributos e estudos que servem de suporte ao governo do estado no gerenciamento eficiente das viaturas.

Este trabalho, portanto, objetiva analisar o consumo de combustível destas viaturas utilizando do processo de mineração de dados para criação de uma ferramenta de suporte a previsão, com o fim de auxiliar no controle interno.

2.2 Mineração de Dados

Mineração de Dados é uma área da ciência da computação que ficou evidente nos últimos 20 anos, devido, principalmente, à popularização da internet e à grande adesão dos computadores pessoais e dispositivos portáteis integrados à internet. O volume de dados gerados por todos estes equipamentos e armazenados em nuvem, necessitou de técnicas sofisticadas para seu tratamento.

Diversas técnicas podem ser aplicadas para compor informação e conhecimento a partir de dados. Entretanto, segundo afirma Amaral (2016), o volume de dados gerados nos últimos anos carecia de técnicas de análise sofisticadas e com alto poder computacional. Desta forma, entende-se que apenas com mineração de dados seria possível extrair, além de informação e conhecimento, *insights* e sabedoria para negócios e organizações.

As principais tarefas executadas a partir da mineração de dados consistem em classificar, prever (regressão), agrupar, associar e detectar anomalias.

De acordo com a pesquisa desenvolvida, busca-se demonstrar a eficácia ao se utilizar técnicas de mineração de dados, com foco em tarefas de previsão, para, a partir de dados discretos, informar ao usuário final o comportamento de consumo de combustível de um veículo no futuro. Esse resultado é possível, devido ao processamento de dados defasados.

Na próxima seção serão apresentadas as técnicas de mineração utilizadas neste trabalho.

2.2.2 Estimação com Algoritmo de Regressão

Regressão é uma técnica estatística que busca prever eventos futuros de um conjunto de dados a partir do comportamento, no tempo, destes mesmos dados. Ela é possível ser aplicada a partir da verificação de correlações fortes entre as variáveis presentes.

A correlação é o que indica a possibilidade de se construir uma regressão. Sua execução busca explicar o comportamento de uma variável a partir do comportamento de outra. Havendo, neste contexto, correlações fortes e fracas, positivas e negativas.

As correlações fortes apresentam valores próximos a 1, portanto, quando mais distante de 1, as correlações entre as variáveis são fragilizadas. Uma vez que o resultado indique que os dados se comportam na mesma direção, as correlações são chamadas de positivas. Por outro lado, se os dados apresentam comportamento inverso, ou seja, quando X aumenta Y diminui, a correlação pode ser forte mas negativa.

Esta medida estatística é basilar para o cálculo e compreensão da regressão, seja ela linear, múltipla, ou logística. As variáveis presentes no conjunto de dados necessitam de correlação significativa para que possam se prever.

Na classificação o valor que se busca prever é a classe. Mas, na regressão, como aplicamos a técnica a valores numéricos, chamamos de variável dependente. A variável dependente é aquela que buscamos prever a partir da aplicação das funções matemáticas e do comportamento da variável independente.

A regressão linear faz parte de um conjunto de modelos baseados em "métodos estatísticos capazes de modelar a relação entre uma variável dependente e um ou mais variáveis independentes" (CASTRO, 2016). Nesta perspectiva, é possível compreender que, através da aproximação dos dados de treinamento pela equação de regressão, torna-se possível variável.

Neste trabalho, a aplicação da regressão permitiu a previsão do comportamento do consumo de combustível, com foco na geração de informação

para melhor tomada de decisão dos envolvidos na gestão da frota de veículos. Para tanto, foi necessária a aplicação do modelo SARIMA, que além da regressão, considera em seus cálculos a sazonalidade, a integração e a média móvel. O modelo SARIMA é discutido na próxima seção.

2.2.3 Previsão com SARIMA

O modelo SARIMA (*Seasonal Autoregressive Integrated Moving Average*) é utilizado com o objetivo de, a partir de uma série temporal, prever o comportamento de uma variável. Este modelo faz parte do grupo de técnicas que trabalham com séries temporais que apresentam sazonalidade, para predição de dados.

Assim, destaca-se a necessidade de que o conjunto de dados apresente frequência adequada de registros, por exemplo, em horas, dias, semanas, meses ou anos. No qual, a ausência de registros pode resultar em um modelo com baixa acurácia ou resultados inconsistentes com a realidade. Este modelo pode ser aplicado, dada suas características, em previsões de dados sociais, econômicos, de engenharia, entre outros problemas, como afirma KHASHEI (2012).

Seu funcionamento parte da compreensão do ARIMA, um modelo, também preditivo, mas que pode ser aplicado a dados sazonais ou não sazonais (MORETTIN & TOLOI, 2006). O ARIMA é o conjunto formado pela auto regressão (AR), a parte integrada (I), e pela média móvel (MA).

A **auto regressão** sazonal infere que o valor da previsão depende dos seus valores defasados, por exemplo, um mês X será previsto no ano A com base no mesmo valor do mês X para seus anos anteriores. A parte **integrada** ordena a diferenciação, que é um recurso estatístico que busca, através de transformações, como logarítmicas, estabilizar a série temporal. Ou seja, estabilizar a média, reduzindo a tendência e a sazonalidade. A **média móvel** utiliza valores de erros de previsões anteriores para indicar que o erro é uma combinação linear dos termos dos erros.

Neste contexto, o SARIMA representa a aplicação do ARIMA em conjunto de dados de série temporal que apresentam características de sazonalidade (Seasonal). Sua aplicação prevê um treinamento a partir de quatro passos, com ciclos iterativos, segundo KHASHEI (2012) *apud* BOX and JENKINS (1976):

- a) Identificar a estrutura do SARIMA (valores sazonais e não sazonais);
- b) Estimar parâmetros desconhecidos;
- c) Realize testes de adequação na estimativa resíduos;
- d) Preveja resultados futuros com base nos dados conhecidos.

O SARIMA foi aplicado neste trabalho devido a sua característica não-estacionária – quando os dados em sua média e variância dependem do tempo – e características do próprio conjunto de dados. O conjunto de dados apresenta um comportamento temporal, ou seja, a execução do abastecimento de combustível, o modelo do carro, e o próprio tempo, em dias.

2.3 Trabalhos Relacionados

A literatura aplicada sobre o tema deste trabalho é tímida, destacando os trabalhos abaixo sobre análises do consumo de combustível.

O primeiro *paper*, desenvolvido por WANG *et al* (2008), *Modelling of the fuel consumption for passenger cars regarding driving characteristics* explora os padrões de consumo de combustível. Sua questão de pesquisa está relacionada com a influência dos padrões de direção no consumo de combustível.

O foco deste trabalho é a utilização de um *portable emissions measurement system* (PEMS) desenvolvido pelo *Department of Environmental Science and Engineering of Tsinghua University*. Este sistema coletou dados de consumo e emissão de combustível nos veículos. Os resultados mostram forte correlação entre velocidade do carro e consumo de combustível, além de um aumento considerável no consumo quando o carro é acelerado.

Em 2014, Niebel *et al*, escreveu o *paper Telematics for the Analysis of Vehicle Fleet Fuel Consumption*. O trabalho demonstra uma verificação sobre o consumo de combustível de uma frota de veículos, aplicada à realidade de pequenas e médias empresas. O objetivo é o aprimoramento da inteligência de negócio aplicada à gestão de frotas, entregando estatísticas em tempo real.

No estudo, a análise faz uso dos dados capturados, são eles: a) o tipo do veículo; b) o estado do motor; c) e o estado do carro. Os autores do *paper*, buscam compreender, o estado do carro, sendo possível encontrá-lo como parado, em movimento, ou parado, mas com o motor ligado. Como indicadores de desempenho,

foram selecionados a quilometragem e o número de paradas.

Neste trabalho, é destacado que toda a modelagem, inclusive o tipo de combustível e o desempenho desejado, foram aplicados para os padrões da Europa, de acordo com normas técnicas.

A pesquisa fez parte de um projeto chamado de *FleetAnalytcs*, simplificado para o *SimpleFleet*, no qual o objetivo é tornar acessível o gerenciamento de frotas para pequenas e médias empresas.

Por fim, em 2020, Gardner *et al*, escreveram *Driving with Data in the Motor City: Mining and Modeling Vehicle Fleet Maintenance Data*.

O trabalho descreveu sobre as manutenções realizadas em frota de viaturas da cidade de Detroit, do estado do Michigan, nos EUA. Os autores destacam que os gastos com a frota são de 7.7 milhões de dólares por ano apenas com manutenção.

O *paper* é um estudo de caso que fornece um *benchmark* baseado em dados, demonstrando um conjunto de métodos para auxiliar na compreensão e previsão de dados para grandes conjuntos de dados de manutenção de veículos.

Apresenta análises para abordar três questões-chave levantadas pelas partes interessadas, relacionadas para descobrir padrões de manutenção multivariados ao longo do tempo, previsão de manutenção; e previsão de nível de veículo e custos de frota.

No contexto do trabalho em tela, os trabalhos de Wang (2008), Niebel (2014) e Gardner (2020), corroboram para a importância de elementos que impactam no consumo de combustível, na gestão de frotas, na análise do comportamento do condutor, e na efetividade da manutenção dos veículos que compõem a frota. Sendo estes, elementos de destaque que podem reduzir o consumo de combustível.

Nesta perspectiva, a partir do escopo deste trabalho, compreende-se a importância de um modelo preditivo que auxilie a PM-PE na gestão da frota e previsão do consumo. Assunto que será abordado nas seções posteriores.

3 Materiais e Métodos

3.1 Descrição da base de dados

Mensalmente, relatórios de resultado de consumo e gastos com combustível das 82 unidades de gestão são enviados para o setor de gestão de combustível. Este setor realiza um parecer que fica a disposição da CPA para análise de não conformidades. A base de dados da PM-PE correspondente à frota de veículos no período de 2018 à setembro de 2020 contém cerca de 700 (setecentas) mil entradas e vinte atributos. Os atributos da base de dados estão descritos na Tabela 1.

Tabela 1: Descrição da base de dados.

Atributo	Descrição
Autorização	Gerado pelo Sistema e único para cada evento de abastecimento
Hodômetro	Marcação de KM no momento do abastecimento
Nome Fantasia	Nome de fantasia do posto de combustível
Cidade	Cidade de Abastecimento
UF	UF de abastecimento
Serviço	Combustível abastecido
Quantidade	Total de litros abastecido
Unitário	Valor por litro
Valor	Total da despesa
DataHora Trans	Data e hora da transação
Num cartão	Número do cartão de abastecimento
Tipo Veículo	Tipo do veículo abastecido
Modelo Veículo	Modelo do Veículo
Deslocamento	KM percorrido com base no hodômetro
Consumo	Deslocamento/Quantidade

Condutor	Motorista
Veículo	Veículo utilizado
Centro de custo	setor da PM responsável pelo veícul
Ano FAB	Ano de fabricação
Ano MOD	Ano do modelo

Em uma análise preliminar, constatou-se que a base de dados da frota da PMPE não é composta por apenas informações relativas ao consumo de combustível, pois, por exemplo, existem outros tipos de serviços incluídos. Sendo assim, se faz necessário uma análise mais detalhada dos dados, objetivando um entendimento mais detalhado do seu conteúdo, e um posterior pré-processamento.

3.2 Análise Descritiva dos dados

A base de dados da PMPE é formada por entradas que descrevem a execução de onze tipos diferentes serviços: de 173 modelos diferentes de veículos de 27 marcas; por 4.373 veículos; por 9.466 condutores; de 97 Centros de Custo (batalhões); e em 187 cidades. Constatou-se, também, um desbalanço na quantidade de entradas entre veículos e a existência da presença de *gaps* (espaços vazios ou registros ausentes) na sequência temporal de registros.

Foram escolhidos dois atributos dos dados, sendo eles: 1. o consumo de combustível; 2. a marca do carro. O primeiro sendo numérico, e o segundo nominal. Após uma etapa de pré-processamento foram feitas análises de estatística descritiva, tais como: média, variância, máximo, mínimo e visualização gráfica.

O valor do Consumo de Combustível é calculado como a razão entre o deslocamento do veículo (Km) e a quantidade de combustível abastecido (l), sendo medido em quilômetros por litro. Este atributo é de suma importância tanto para a CPA quanto para a PM-PE e devem apresentar o mínimo de anomalias possível.

A Tabela 2 apresenta os limites inferior e superior e o ponto médio de todas as entradas do atributo Consumo. Com base nessas informações, é possível perceber que há alguma anomalia, pois dificilmente uma viatura conseguirá ter um consumo de combustível de 0,05 Km/l. O ponto médio é um valor aceitável dada as diversas

características e especificidades operacionais natas desse tipo de veículo (por exemplo: gaiola, ocorrências, transporte de animais, trânsito, etc).

Tabela 2: Medidas de distribuição de frequência dos dados de consumo: limites inferior e superior e média.

Medida	Valor
Limite Inferior	0.05
Limite Superior	23.21
Média	08.02

A Figura 1 mostra as frequências absoluta e relativa de determinados intervalos de consumos de combustível. Por meio dela é possível perceber como e onde os valores de consumo de combustível estão concentrados. Os intervalos 0 e 1 representam, muito provavelmente, a proporção de entradas contendo anomalias na base de dados, pois é muito difícil que haja algum veículo consumindo tanto combustível. Já o intervalo 3 [6.95 - 9.26] mostra onde encontra-se a preponderância de valores. A distribuição de frequência indica, também, a necessidade ou não da execução de processos mais específicos de limpeza/tratamento de dados.

	intervalos	frequencia_absoluta	frequencia_relativa
0	0 -> 2.316	298	0.000514
1	2.32 -> 4.63	15034	0.025924
2	4.63 -> 6.95	163293	0.281576
3	6.95 -> 9.26	263263	0.453960
4	9.26 -> 11.58	113201	0.195199
5	11.58 -> 13.90	19022	0.032801
6	13.90 -> 16.21	3611	0.006227
7	16.21 -> 18.53	1267	0.002185
8	18.53 -> 20.84	599	0.001033
9	20.84 -> 23.16	337	0.000581

Figura 1: Distribuição de frequência para dados de consumo.

A Figura 2 mostra o histograma do Consumo de Combustível. Por meio dele é possível constatar que a grande maioria dos valores de consumo de combustível das entradas da base de dados encontram-se situados entre ~7 e 9 Km/l.

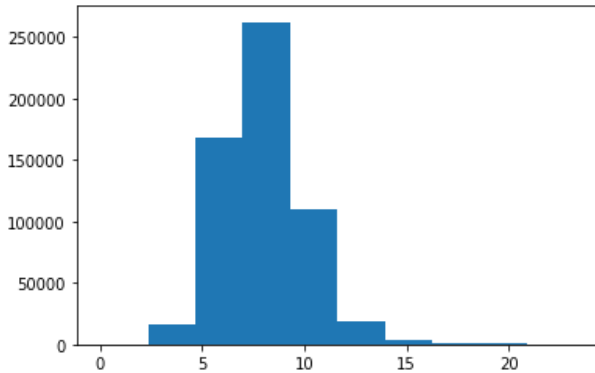


Figura 2: Histograma do consumo.

A Figura 3 apresenta o gráfico *BloxPlot* do Consumo de Combustível. É possível verificar de forma gráfica um resumo estatístico dos valores do consumo de combustível: a mediana é de aproximadamente 8 Km/l; o quartil inferior é próximo de 6 Km/l; o quartil superior é cerca de 9 Km/l; o limite superior é ~ 14 km/l; e o limite inferior está situado próximo à 3 Km/l. Há uma quantidade considerável de entradas na base de dados contendo valores acima de 14 km/l e abaixo de 3 Km/l que devem ser investigados. A distância entre a mediana e as outras medidas indicam um desvio padrão alto. O desvio padrão alto pode estar relacionado com a análise conjunta de diferentes modelos de veículos, regiões geográficas e características do trânsito.

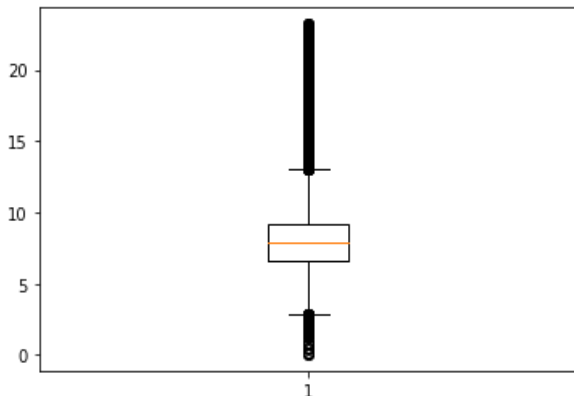


Figura 3: Gráfico BoxPlot dos dados de Consumo .

A frota de veículos da PM-PE é diversa. Ela é formada por uma variedade de tipos de veículos, por exemplo carro, caminhonete, ônibus, caminhão, trator, ambulância, motocicleta, etc, que são aplicados em regiões e operações específicas. Os veículos são oriundos de montadoras de automóveis diferentes, com diversos modelos diferentes. A escolha do

modelo e marca é estabelecida por meio de licitação pública.

Após o pré processamento dos dados foram mantidos apenas as entradas referentes aos veículos do tipo carro e caminhonete.

A Figura 4 mostra as frequências absoluta e relativa dos modelos dos veículos selecionados para análise. Como os modelos são muito diversos, separamos os modelos com mais entradas, ou seja, os com o número de frequência relativa de registros de consumo maior que 0.05, dividindo assim os modelos entre SPACEFOX, SPIN, HILUX, S 10 e OUTROS, onde a última categoria agrupa todos os modelos não mencionados anteriormente.

A Figura 5 mostra o histograma dos Modelos de Veículo. Por meio dele é possível constatar que a grande maioria das entradas da base de dados é dos veículos do modelo SPIN, seguido do SPACEFOX, e dentro das caminhonetes, a S 10 e a HILUX são as grandes representantes.

	Modelos	Freq. Absoluta	Freq. Relativa
0	SPACEFOX	137562	0.235715
1	SPIN	243492	0.417228
2	HILUX	56334	0.096529
3	S 10	32914	0.056399
4	OUTROS	113292	0.194128

Figura 4: Distribuição de frequência para dados de modelo de veículo.

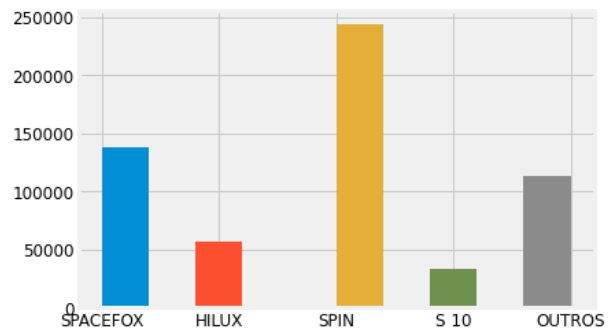


Figura 5: Histograma dos modelos dos automóveis.

3.3 Pré-processamento dos Dados

Durante a avaliação dos dados foram observadas inconsistências nos dados. Buscando saná-las, foram executadas três etapas de pré-processamento. Sendo elas:

Limpeza dos dados, onde foram removidos valores que não estavam presentes em todos os casos (por exemplo nas datas dos serviços, em algumas casos havia apenas o dia, mês e ano e em outros havia também a hora, e para padronizar foram removidas todas as horas das datas). E também foram re-avaliados dados inconsistentes (por exemplo o consumo dos veículos vezes aparecia como negativo ou na ordem de milhares de quilômetros por litro), para fazer isso recalculamos o consumo de acordo com a distância movimentada pelo carro (usando o hodômetro) e a quantidade de litros abastecida no período desse deslocamento. E além disso foram removidos os *outliers* utilizando a média mais ou menos três vezes o desvio padrão.

Redução, foram removidos tipos de veículos e de serviços que fogem do objetivo dessa análise, por exemplo, barcos, caminhões, troca de óleo, manutenções etc.

Transformação, alguns dados tiveram sua formatação padronizada, sendo eles a data, serviços e modelos de carro.

3.4 Metodologia Experimental

Como o comportamento de consumo das viaturas pode ser descrito por uma série temporal e dada a sazonalidade descrita pelos *stakeholders* e verificada nos dados (gráfico 2), foi utilizado o algoritmo SARIMA e a sua implementação foi facilitada com o uso do pacote *Stats Model* do *Python* (StatsModels). Como o SARIMA trabalha com séries temporais, foi extraída uma série temporal de consumo de combustível dos dados separados por modelo de carro. Para dias que tinham mais de um valor foi calculada a média do dia para aquele modelo. O SARIMA retorna inicialmente uma previsão dentro dos dados de teste, onde foi possível comparar as previsões com os dados reais e assim calcular a taxa de erro.

4 Análise e Discussão dos Resultados

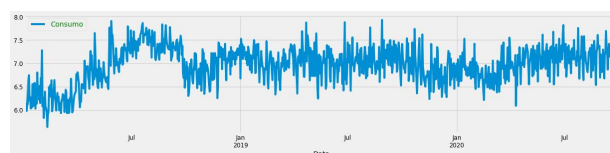
4.1 Resultados

A aplicação do algoritmo SARIMA, no conjunto dos dados de série temporal do comportamento do consumo de combustível pela frota de carros da PM-PE, apresentou resultado satisfatório na ordem de 84% de acertos.

Buscou-se verificar no resultado a correta previsão dos dados, utilizando a base seccionada em treino e teste. O conjunto com dados de treino correspondeu aos valores de consumo de combustível diário por tipo de veículo, no período de 2018 a 2019. Por outro lado, o conjunto de teste compôs os mesmos dados para o ano de 2020.

O gráfico 1, representa o comportamento do consumo médio diário dos dados no período de janeiro de 2018 a setembro de 2020.

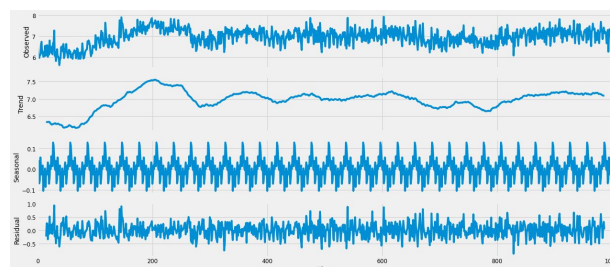
Gráfico 1 - Comportamento da média de consumo de combustível



Após a aplicação do algoritmo SARIMA, o resultado de viabilidade é apresentado no gráfico 2, em que é possível verificar quatro comportamentos.

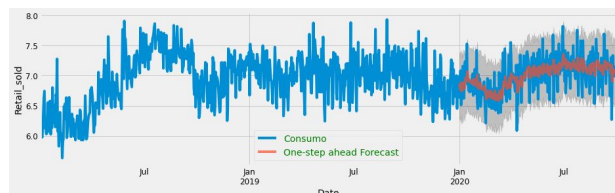
O primeiro comportamento (raia) é similar ao gráfico 1, demonstrando o comportamento da série. Comparativamente, ainda no gráfico 2, são demonstrados a tendência do comportamento, que corrobora com o comportamento real dos dados; a sazonalidade, que representa os picos e vales que se repetem dado um evento; e por fim o ruído.

Gráfico 2 - Resultados de viabilidades do algoritmo SARIMA



Verificado que o conjunto de dados estava alinhado com o objetivo do algoritmo, o passo seguinte resultou na execução da previsão.

Gráfico 3 - Previsão do consumo de combustível no meses do ano de 2020 (Janeiro à Setembro)



O resultado é dado na forma de um intervalo de valores de consumo. Foram comparados os valores reais com os valores previstos, e houve uma taxa de acerto de 84%.

Este último gráfico, demonstra que os valores treinados conseguem prever a linha central que corrobora com o comportamento dos dados. Notadamente, existe um intervalo de segurança apresentada pelo algoritmo, que apresenta um intervalo máximo e mínimo possível dos dados, e a previsão, se torna a linha central deste intervalo.

4.2 Discussão

Como mencionado acima, os trabalhos de Wang (2008), Niebel (2014) e Gardner (2020) mostram aspectos da gestão de frotas de veículos responsáveis pela redução do consumo de combustível e, em consequência, do valor destinado à compra de tais insumos. A PM-PE, como um órgão estatal, deve versar pela otimização do emprego dos recursos públicos nos custos operacionais da corporação. O valor destinado à compra de combustível corresponde a uma fatia significativa do orçamento. Sendo assim, seja na literatura ou na PM-PE, há uma busca por modelos que auxiliem no racionamento dos gastos designados à frota de veículos.

O registro do consumo de combustível de uma viatura ou de uma frota de viaturas, ao longo do tempo e dividido em épocas, gera uma série temporal sazonal. O algoritmo SARIMA foi desenvolvido com o objetivo de prever o comportamento de séries temporais sazonais com base nos registros pregressos. A acurácia do SARIMA é sensível à quantidade de registros e, também, da presença de *gaps* na série temporal, isto é, quanto mais registros e menos *gaps* a

probabilidade de se obter uma melhor acurácia aumenta.

Apesar da base de dados da PM-PE compreender apenas dados de 2018 até setembro de 2020, conter *gaps* e um desbalanço na quantidade de registro por viaturas, os resultados obtidos por meio da aplicação do SARIMA na variável CONSUMO de um dos modelos de viaturas da PM-PE mostraram uma acurácia de 84% e indicam que é possível utilizá-los para, *a priori*, prever o consumo de combustível das viaturas para **momentos futuros**, facilitando o planejamento do orçamento da corporação. Além, ao aplicar a previsão à um **momento corrente**, os resultados do SARIMA podem ajudar os auditores na identificação de não conformidades, pois é possível identificar veículos e/ou modelos de veículos que consomem ou estão consumindo combustível além dos valores mínimos aceitáveis e diminuir o espaço amostral da busca e focar apenas nos resultados que estão fora da normalidade.

Foram testados outros cenários de previsão, modificando o modelo de viatura e/ou filtrando por uma viatura específica, porém a resposta do SARIMA não foi satisfatória. O limitante foram a quantidade de registros e a presença de *gaps*, espaços sem registros de comportamento por uma viatura. Espera-se que a acurácia do modelo proposto seja maior e a implementação de filtros mais específicos (por veículos, centro custo, etc) seja possível, caso disponha-se de uma base de dados contendo uma série temporal mais longa (mais períodos sazonais), com uma maior regularidade dos registros históricos, sem *gaps* e melhor balanceada.

5 Conclusões e Trabalhos Futuros

A perspectiva analítica que busca aplicar mineração de dados a problemas de previsão no consumo de combustível de frotas, merece atenção especial devido às especificidades de cada conjunto de dados.

No contexto e escopo da PM-PE se compreende a importância, por exemplo, de manter o sigilo, a anonimização, entre outros elementos com foco na segurança dos atores finais envolvidos no processo, e na própria atuação que exige um grau de segurança específico. Além, destes fatores, a própria configuração dos veículos,

compostos por equipamentos extras, alteram fatores que podem influenciar no consumo.

Diante destas especificidades, buscou-se encontrar as melhores correlações entre os atributos, de modo que a previsão de consumo seja realizada de forma positiva, e auxilie a tomada de decisão assertiva.

Este trabalho também propõe continuidade. De modo que, em seu teor prático, sejam agregadas novas métricas, a partir da integração com bases de dados de fabricação dos veículos, do mapa de relevo da cidade do Recife e de gestão do fluxo de trânsito. Também é interessante validar com outros algoritmos de previsão por séries temporais com sazonalidade. Nesta perspectiva, buscando um grau mais profundo de assertividade, não contemplado neste trabalho, devido ao limitador tempo.

Referências

- AMARAL, Fernando. Aprenda Mineração de Dados: teoria e prática. Rio de Janeiro, RJ : Alta Books, 2016.
- BRESSER-PEREIRA, Luiz Carlos (2008). "O modelo estrutural de gerência pública." Revista de Administração Pública – RAP, Rio de Janeiro 42(2) : 391 – 410, Março/Junho.
- CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. Introdução à Mineração de Dados: conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.
- GARDNER, Josh et al. Driving with Data in the Motor City: Mining and Modeling Vehicle Fleet Maintenance Data. arXiv preprint arXiv:2002.10010, 2020.
- KHASHEI, Mehdi; BIJARI, Mehdi; HEJAZI, Seyed Reza. Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting. Soft computing, v. 16, n. 6, p. 1091-1105, 2012.
- MARIZ, Filipe Mendes. Avaliação e comparação de versões modificadas do algoritmo KNN. 2017. 54 f. TCC (Graduação) - Curso de Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife, 2017. Disponível em: <https://www.cin.ufpe.br/~tg/2017-2/fmm4-tg.pdf> f. Acesso em: 01 nov. 2020.
- MORETTIN, P.A . & TOLOI, C.M.C. Análise de Séries Temporais. São Paulo, Edgard Blücher, 2006.
- NIEBEL, Wolfgang et al. Telematics for the Analysis of Vehicle Fleet Fuel Consumption. In: International Conference on Transport Systems Telematics. Springer, Berlin, Heidelberg, 2014. p. 461-468.
- TRIBUNAL DE CONTAS. Manual de Auditoria e de Procedimentos, vol. I e II. Ed., Tribunal de Contas, 1999.
- WANG, Haikun et al. Modelling of the fuel consumption for passenger cars regarding driving characteristics. Transportation Research Part D: Transport and Environment, v. 13, n. 7, p. 479-482, 2008.
- WEBER, Max (1971). "Os fundamentos da organização burocrática: uma construção do tipo ideal." In: CAMPOS, Edmundo. Sociologia da burocracia. Rio de Janeiro: Zahar.
- BROWNLEE, JASON. How to Remove Outliers for Machine Learning. Machine Learning Mastery, 2018. Disponível em <https://machinelearningmastery.com/how-to-use-statistics-to-identify-outliers-in-data/> Acesso em 27 de Outubro de 2020.
- StatsModels. Disponível em <https://www.statsmodels.org/stable/index.html> > Acesso em 27 de Outubro de 2020.