RMP Revista dos Mestrados Profissionais – RMP
HTTPS://PERIODICOS.UFPE.BR/REVISTAS/RMP
ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

Uso de Mineração de Dados para Identificação de Atos Suspeitos na Prova de Direção da CNH

Letícia Toledo Maia Zoby - letmaia@gmail.com[1]
Gabriel Lima Gomes - gabriel.lg08@gmail.com[2]

Resumo - Uma das principais causas de perda de receita é a fraude. E com o crescente volume e armazenamento de dados várias empresas torna-se inviável a análise manual de todos os dados. O uso do KDD é (*Knowledge Discovery in Database* - Descoberta de conhecimento em base de dados) é um dos recursos no auxílio de gestão de grande quantidade de dados e transformá-los em informações. Este trabalho descreve uma das etapas do KDD, a mineração de dados, para extrair informações sobre índice de aprovação de exame de direção para obtenção da CNH, de uma base de dados real, para identificar possíveis casos de fraudes nesse exame. Foi utilizado detecção de *outlier* utilizando o *boxplot* e *cluster* com algoritmo k-means. Através dessas aplicações foi possível identificar os índices de aprovação atípico e extrair informações sobre esses casos encontrados.

Palavras-chave: mineração de dados, CNH, boxplot, k-means.

Datamining to Identify Suspicious Acts in the CNH Driving Test

Abstract - One of the main causes of lost revenue on firms is fraud and manual analysis to prevent losses nowadays it is inviable becauses increasing volume and storage of data fastly. In parell, the use of KDD is (Knowledge Discovery in Database)is one of the resources in the aid of managing large amounts of data and transforming them into information. This paper describes one of the KDD's steps, data mining, to extract information about the management examination approval index for obtaining the CNH, from a real database, aims to identify possible cases of fraud in that examination. Outliers detection using boxplot and cluster with k-means algorithm was used. Through these applications it was possible to identify the atypical approval rates and extract information about these cases found.

Keywords: data mining, CNH, boxplot, k-means.

Data de Submissão: 24/04/2020
Data de Aceitação: 29/04/2020

[1] Centro Universitário IESB - Brasília - Engenharia de Sistemas Eletrônicos e de Automação, [2] Centro Universitário IESB - Brasília - Analista de dados

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

1. Introdução

O uso da tecnologia é empregado em várias empresas como sinônimo de desempenho favorável, garantia de melhores resultados, aumento na produtividade e no lucro, melhoria nas tomadas de decisões e no combate a fraudes. Detectar fraudes é importante nas empresas uma vez que as perdas geradas representam um fator negativo.

Devido à grande quantidade de dados coletados nos sistemas de informação durante anos, a utilização do recurso de descoberta de conhecimento em base de dados tem ajudado na seleção de informações úteis para as companhias [Hekima, 2014].

Existe um grande esforço dos departamentos de trânsito do Brasil (DETRANs) para manter seus sistemas robustos e seguros, de modo a evitar fraudes nos processos para a emissão da Carteira Nacional de Habilitação (CNH). Infelizmente, há sempre manchetes de jornal mostrando ações fraudulentas, como em fevereiro de 2016 foi noticiado no site Correio Braziliense "Quadrilha cobrava até R$ 6 mil para fraudar exames do Detran-DF" [Stacciarine, Saraiva and Cardim 2016].

O processo para obtenção da CNH visa capacitar o candidato com aulas teóricas, estudando conteúdos como, direção defensiva, legislação de trânsito, primeiros socorros, entre outras e com formação prática de direção veicular nos Centros de Formação de Condutores (CFC). A habilitação é importante para impedir que pessoas não aptas, como, pessoas com algum problema de visão que não há correção, conduzam veículos resguardando sua vida e a do próximo.

A credibilidade dos DETRANs pode ser enfraquecida perante a sociedade com um escândalo, além de um sentimento de injustiça quando se sabe que muitos candidatos obtiveram a CNH sem as devidas exigências cumpridas, enquanto os demais cumpriram. Assim, os órgãos competentes procuram modernizar esse processo procurando evitar atos fraudulentos, como, sistema de biometria, acesso mais restritos aos sistemas, armazenamento de logs dos sistemas, câmeras de monitoramento.

O processo de Descoberta de Conhecimento em Banco de Dados (KDD - *Knowledge Discovery in Database*) preocupa se com o desenvolvimento de métodos e técnicas para extrair conhecimento a partir das informações existentes nos dados. E tem como principal etapa o processo de mineração de dados, consistindo na execução prática de análise de algoritmos específicos que produz uma relação particular de padrões a partir de dados.

Neste artigo, propõe uma metodologia para auxiliar os órgãos competentes a identificarem casos suspeitos no processo da obtenção da CNH utilizando recursos disponíveis no processo de KDD. Foram empregados técnicas de mineração para identificar anomalias (*outlier)* e padrões (*cluster)* em grande base de dados.

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

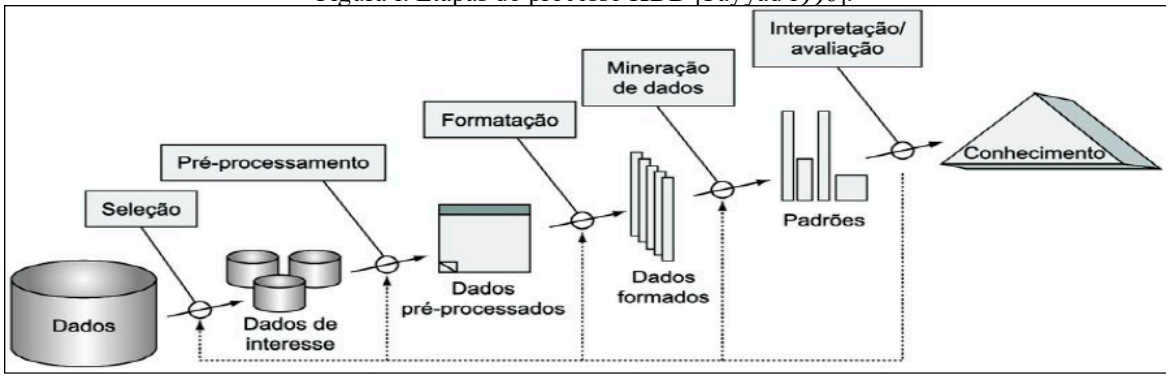
2. Descoberta de Conhecimento em Base de Dados e Mineração de Dados

Com o crescente volume de dados é inviável analisá-los de forma manual. Torna-se necessário utilizar as ferramentas disponíveis para análise de grandes conjuntos de dados.

O KDD é um dos métodos disponíveis para a possibilidade de analisar grandes conjuntos de dados. Ele não trivial para extração de informações/padrões válidos, novos e potencialmente compreensível em grandes conjuntos de dados [Fayyad 1996].

Para extração de conhecimento em grandes bases de dados, envolve muitas etapas que vão desde manipulação e recuperação de dados, à inferência matemática e estatística, pesquisa e raciocínio [Fayyad 1996]. Na Figura 1 é possível observar o processo de KDD com todas as etapas existentes.

Figura 1. Etapas do processo KDD [Fayyad 1996].



Em seguida será detalhada cada etapa do processo KDD, conforme a Figura 1.

Seleção: Etapa a qual se deve selecionar o conjunto de dados, ou focar em um subconjunto de variáveis ou amostra de dados os quais são para ser descoberto os padrões [Fayyad 1996].

Pré-processamento: nesta etapa ocorre a remoção de ruídos ou valores atípicos e se necessário, coletar as informações essenciais para o modelo ou por conta dos ruídos, e decidir sobre estratégia por falta de campos de dados [Fayyad 1996].

Transformação: esta etapa trata de encontrar recursos úteis para representar os dados, dependendo do objetivo da tarefa, pode ser usando redução de dimensionamento ou métodos de transformação para reduzir o número de variáveis [Fayyad 1996]. Um exemplo de transformação de dados, é um atributo sexo, o armazenamento desse atributo ébastantediversificado nas bases de dados, como, M/F, MASC/FEM, 0/1 etc. Então, esse atributo deverá ser transformado para um único formato.

Mineração de dados: Esta etapa busca por padrões de interesse, em particular de uma forma representativa ou conjunto de tais representações [Fayyad 1996]. Esta etapa será mais detalhada na seção 2.1.

Interpretação: Nessa etapa é realizada a interpretação dos padrões descobertos e a possibilidade de retorno para as etapas anteriores, bem como a visualização dos padrões extraídos, eliminação de redundância ou padrões irrelevantes, e traduzir em termos interessantes e úteis para o usuário [Fayyad 1996].

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

2.1 Mineração de Dados

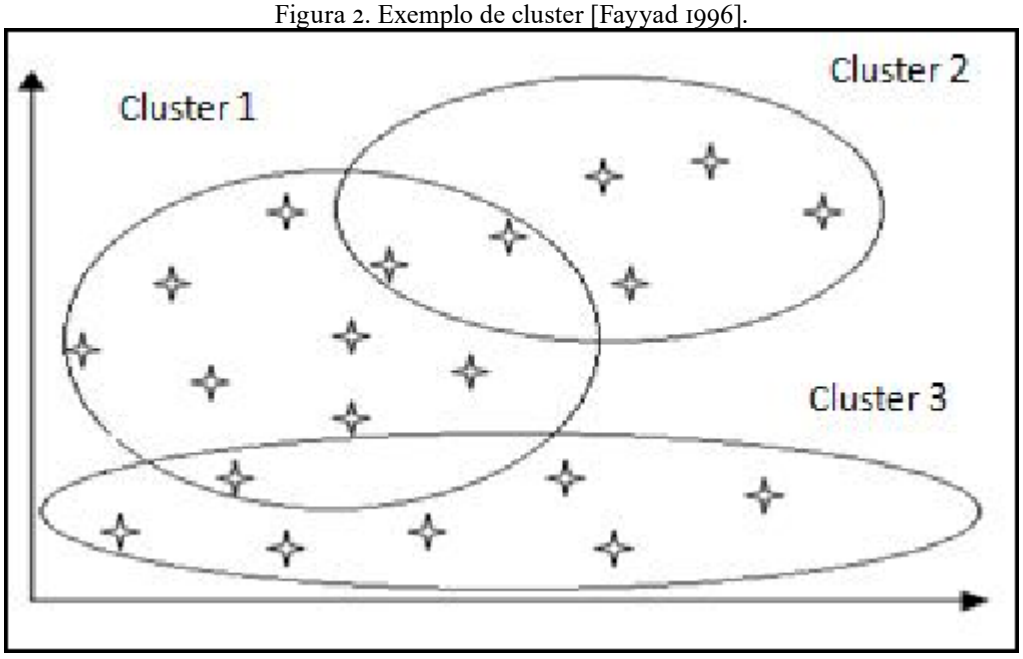
Mineração de Dados - DM (*Data Mining*) é uma etapa do processo do KDD, o qual se aplica algoritmos específicos para extração dos modelos de forma automática ou semiautomática, mas aceitando as limitações da eficiência computacional. Esta etapa é de extrema importância e uma atividade legítima para o processo KDD, desde que se entenda como realizá-la corretamente [Fayyad 1996].

Assim, define-se a tarefa da DM e as técnicas a serem utilizadas para atender os requisitos iniciais que se deseja alcançar. É importante saber distinguir tarefa de técnica de mineração de dados. A tarefa consiste em **do que** está se querendo buscar nos dados, que tipo de regularidades e categorias de padrões temos interesse em encontrar, ou novos tipos de padrões. Dentre as tarefas de mineração têm-se, Análise de regras de associação, classificação e predição, Cluster (agrupamentos) e Outlier (anomalias). A técnica de mineração de dados trata de **como** descobrir os padrões que nos interessam, dentre as principais técnicas, têm-se estatísticas e técnicas de aprendizado de máquina [Amo 2004].

Em seguida será feito uma descrição mais detalhada das tarefas de mineração de dados que foram utilizadas para a execução deste trabalho:

Cluster: mais comum tarefa de descrição, que identifica um finito conjunto de cluster (agrupamento) para descrever os dados. Os *clusters* podem ser mutuamente exclusivos e exaustivos, ou consistem de uma representação rica, como hierárquica ou agrupamentos sobrepostos [Fayyad 1996]. Esta tarefa trabalha sobre dados onde os rótulos de classes não estão definidos. Na Figura 2 é possível observar um exemplo simples de possíveis cluster, os quais, os dados estão em três agrupamentos. É importante observar que os cluster podem sobrepor permitindo que os dados façam parte em mais de um cluster [Han, Kamber and Pei 2012].

Figura 2. Exemplo de cluster [Fayyad 1996].

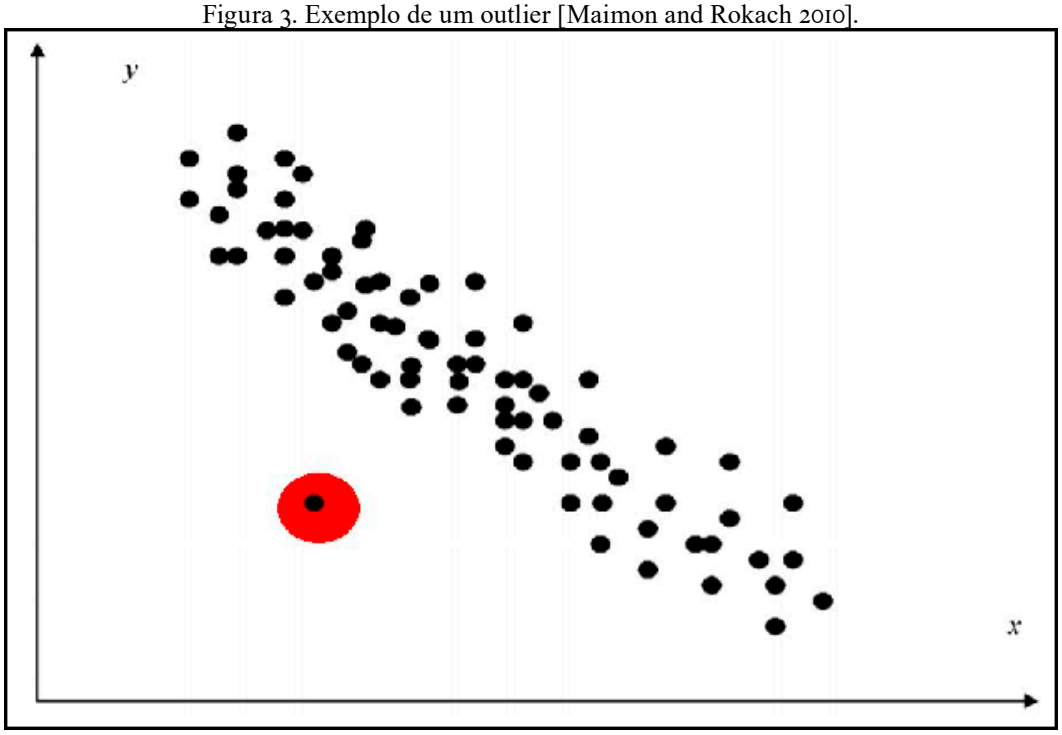


ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

Para obter os padrões da base de dados, existem diferentes algoritmos, dentre eles tem-se o K-Means. Este algoritmo objetiva particionar**n**observações dentre k grupos (cluster), no qual cada observação pertence ao grupo mais próximo da média/centro. O centro de cada cluster é resultante da média de todas as instâncias que pertencem a esse agrupamento [Maimon and Rokach 2010].

Outlier: Consiste em encontrar desvio de padrões, ou seja, dados que não apresentam o comportamento geral da maioria [Amo 2004], descobrir mudanças mais significativas nos dados medidos anteriormente ou valores normais [Fayyad 1996]. Em detecção de fraudes e detecção de falhas esses registros que diferem da maioria, são os que interessam, mas também pode ocorrer que esta situação apenas oculte os principais pontos de um modelo proposto. Assim é interessante identificar esses desvios de valores e eliminá-los antes de finalizar o modelo. *Outliers* é diferente de ruído dos dados. Ruído é um erro aleatório ou desvio em uma medida da variável, em geral estes ruídos não são interessantes para análises dos dados [Han, Kamber and Pei 2012]. Na Figura 3 é possível observar um exemplo simples de detecção de *outlier*.

Figura 3. Exemplo de um outlier [Maimon and Rokach 2010].

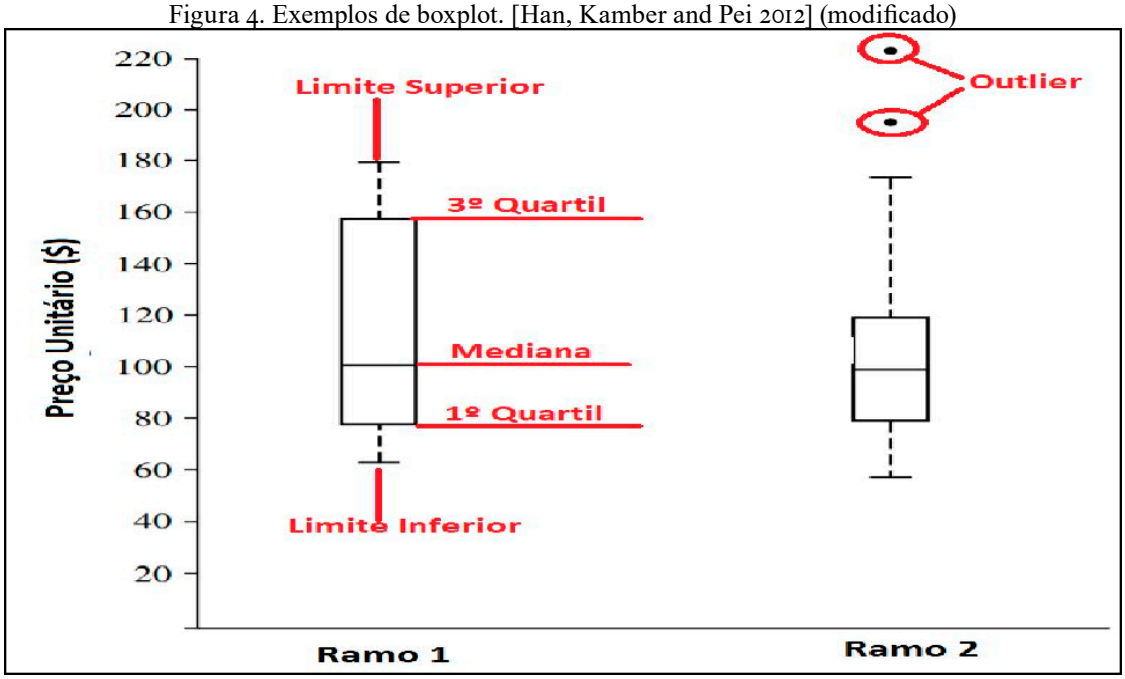


Para realizar a detecção de um *outlier* em uma base de dados, um dos métodos existentes é através de cálculos estatísticos, em seguida será detalhado um dos métodos estatísticos existentes para detecção de valores atípicos.

O *boxplot* é uma das opções existente para visualização de distribuição dos dados. O *boxplot* é composto por cinco valores que são: 1º quartil (Q1), 3º quartil (Q3), mediana, limite inferior e limite superior. A utilização desse recurso é vantajosa por ser uma representação gráfica de fácil análise, quando houver algum ponto fora dos limites inferior ou superior, são considerados *outlier* [Han, Kamber and Pei 2012], conforme a Figura 4.

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

Figura 4. Exemplos de boxplot. [Han, Kamber and Pei 2012] (modificado)



Na Figura 4 são demonstrados dois gráficos *boxplot*, um sem valores discrepantes (à esquerda) e outro com valores discrepantes (à direita). Os valores atípicos são definidos pela Equação 1 [Han, Kamber and Pei 2012]:

$$X < Q1 - 1{,}5 * AIQ \quad (1)$$

Ou pela Equação 2:

$$X > Q1 - 1{,}5 * AIQ \quad (2)$$

Onde, AIQ (Amplitude Interquartil) é definida por Equação 3:

$$AIQ = Q3 - Q1 \quad (3)$$

Com estes gráficos é possível detectar de modo mais rápido e eficaz valores atípicos na base de dados, assim, estes registros terão um tratamento especial para identificar o porquê dessa ocorrência, se houve algum erro nos seus valores ou são possíveis atos de suspeita de irregularidades.

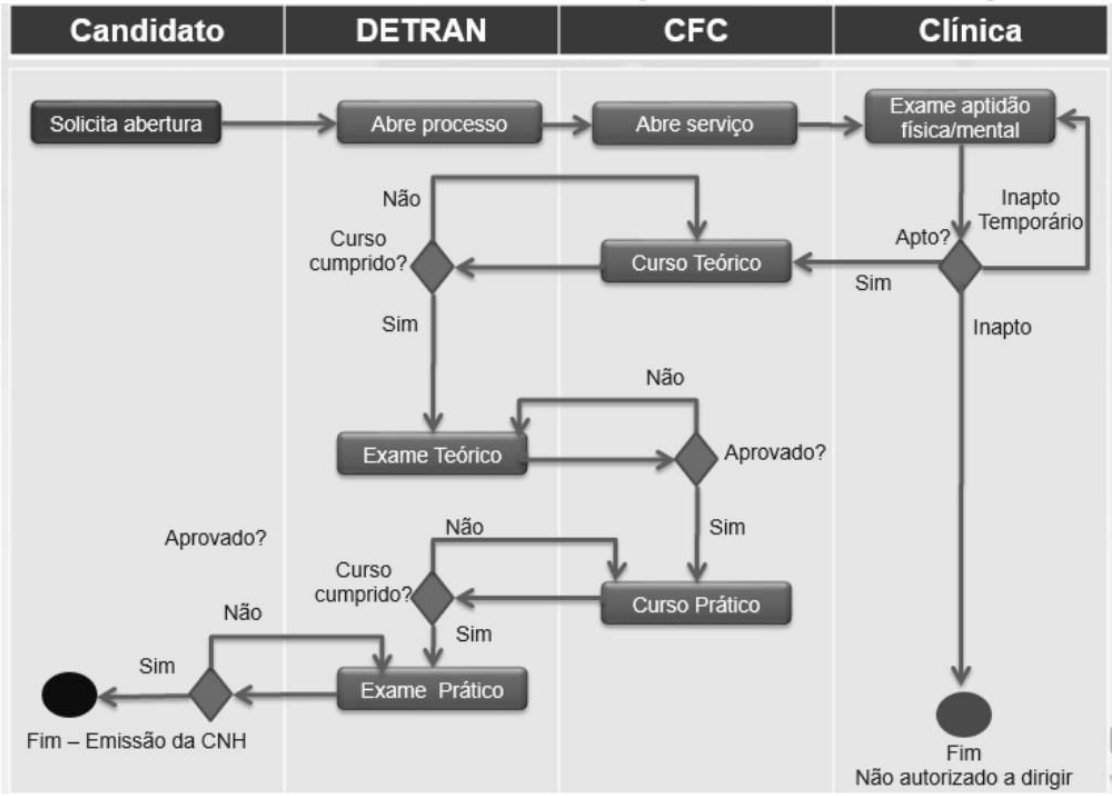
3. Processo de Obtenção da Carteira Nacional de Habilitação

A CNH (Carteira Nacional de Habilitação) é um documento que atesta que a pessoa estáaptaa conduzir um veículo automotor, e o seu porte é obrigatório ao dirigir [Portal Brasil 2014]. Para a obtenção da CNH é necessário preencher alguns requisitos tais como: ser penalmente imputável, saber ler e escrever, e possuir carteira de identidade ou equivalente [Brasil art.1401997].

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

O cidadão que tiver interesse em tirar CNH deverá passar por um processo de formação de condutores até a emissão da habilitação, conforme a Figura 5.

Figura 5. Diagrama do processo para obtenção da CNH.



4. Metodologia

A metodologia para alcançar o objetivo deste trabalho visa estudar o processo de descoberta de conhecimento em base de dados, sobre os casos de fraudes na obtenção da CNH, assim definindo os atributos importantes para compor o modelo para detectar os atos que podem caracterizar alguma irregularidade.

Por questões de sigilo os dados identificatórios como, CPF, CNPJ, nome dos examinadores, instrutores, nome das empresas e códigos de identificação, foram alterados. Também não será divulgado de qual DETRAN é a base de dados usada.

Neste trabalho foram utilizados os seguintes atributos: a) código dos examinadores; b) data da aplicação do exame prático; c) idade; d) sexo; e) estado civil; f) grau de instrução do candidato; g) categoria da CNH; h) se exerce atividade remunerada com a CNH; i) nome do CFC; j) motivo de requerimento do processo; e, l) o resultado final do processo.

Por não haver uma base de dados com fraudes em obtenção da CNH não é possível utilizar técnicas de aprendizado supervisionado, como, classificação. Assim, foram utilizados métodos de aprendizagem não-supervisionado para identificar comportamento anômalos na base de dados, com análise de *outlier* utilizando *boxplot.* Posteriormente, reconhecer padrões

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

nesses casos anômalos utilizando o algoritmo *k-means,* para identificar ações que podem se caracterizar possíveis atos fraudulentos que não são conhecidos.

Para o desenvolvimento deste trabalho foi selecionado uma fase específica da obtenção da CNH para análise, etapa de exame prático. Este exame é realizado pelo candidato perante uma comissão formada por 3 membros, o candidato deverá estar acompanhado por no mínimo 2 (dois) membros da comissão, estes irão avaliar o candidato, assim, aprovando-o ou não. [Gomes and Zoby 2016].

Todas as etapas do KDD foram executadas no desenvolvimento deste trabalho, principalmente visando as etapas de pré-processamento e transformação, para limpar e adaptar os dados. É necessário para uma melhor análise e adequação para a execução do algoritmo *k-means*.

Os softwares escolhidos foram os RStudio[1] para análise de dados e detecção de *outlier*e Weka[2] para a aplicação da mineração de dados.

O RStudio é um programa de desenvolvimento integrado para R, possui versões open-*source*e licenças comerciais disponíveis, funciona em diferentes sistemas operacionais, como, Windows, Linux e Mac. Esse programa foi desenvolvido na linguagem de programaçãoC++, possibilita realizar cálculos estatísticos, científicos de forma fácil e rápida, além de, possibilitar a criação de diferentes gráficos, dentre eles, gráficos para identificação de*outlier* na base de dados [Rstudio 2016].

O WEKA (*Waikato Envronment for Knowledge Analysis*) é um software livre e gratuito desenvolvido pela Universidade de Waikato na Nova Zelândia, implementado na linguagem Java, sua principal característica é a portabilidade podendo ser utilizado em diferentes tipos de sistemas operacionais. Esta suíte contém um conjunto de implementações de algoritmos de diversas técnicas de mineração de dados como: classificação, cluster e associação [Damasceno 2010].

5. Análise e Discussão dos Resultados

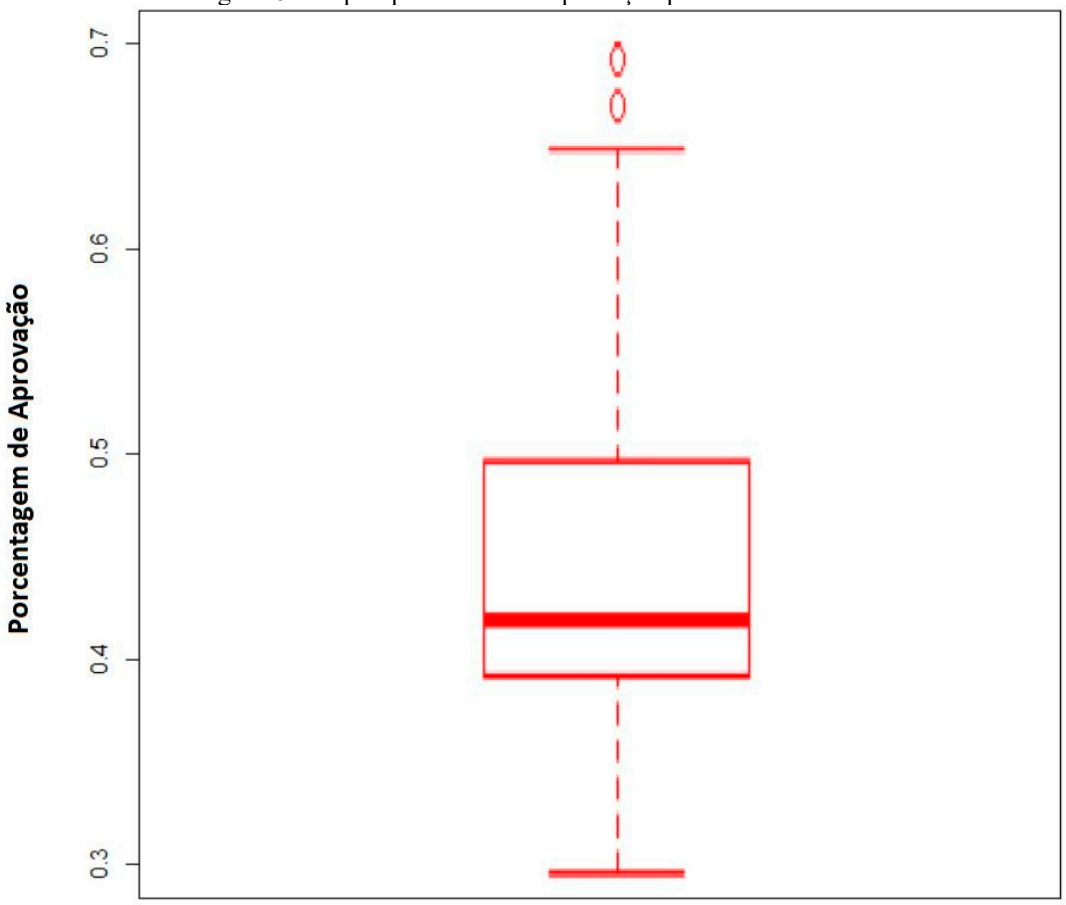
Para identificar *outlier* na base de dados, foi utilizado o *boxplot*, citado anteriormente na tentativa de identificar valores atípicos, assim, podendo identificar quais casos devem ter um foco maior ao aplicar outras tarefas, neste caso o *cluster*, e se é possível identificar possíveis fraudes na realização da prova prática para obtenção da CNH.

1 https://www.rstudio.com/products/rstudio/download/

2 http://www.cs.waikato.ac.nz/ml/weka/downloading.html

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

Figura 6. Boxplot para índice de aprovação por examinadores.



Foram identificados dois valores atípicos, entre 0.65 e 0.70, esses registros são os casos suspeitos, então, realizando uma análise mais aprofundada em seguida.

Tabela 1. Dupla de examinadores identificados como outliers.

CÓDIGO	QTD_EXAMES	QTD_EXAM_AP	PORCENTUAL_AP
EX1520-EX1392	201	137	68,15%
EX1011-EX0465	237	167	70.46%

Por questões de sigilo dos dados identificatórios, os códigos dos examinadores,foramalterados do seu modelo original, já citado anteriormente. As colunas da Tabela 1, **qtd_exames**apresenta a quantidade de exames realizados, **qtd_exam_ap** apresenta a quantidade de exames aprovados e o campo **porcentual_ap** apresenta a porcentagem de aprovação de cada dupla de examinador.

Após a identificação de*outlier*, foi aplicado algoritmo de agrupamento (*cluster*) para tentar identificar os padrões de comportamento desses casos suspeitos, utilizando o software-*Weka*. Para a execução dessa tarefa foi selecionado o algoritmo*k-means*, detalhado na seção 2.1, configurando-o para execução com 4*cluster.*

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

Em seguida serão mostrados e detalhados os resultados obtidos após aplicação do algoritmo para os examinadores.

Tabela 2. Tabela com resultado do cluster para a dupla de examinadores EX1520-EX1392.

CAMPOS	CLUSTER 0 (24)	CLUSTER 1 (15)	CLUSTER 2 (61)	CLUSTER 3 (20)
IDADE	>32	>32	18-25	>32
SEXO	FEMININO	MASCULINO	MASCULINO	FEMININO
ESTADO_CIVIL	SOLTEIRA	CASADO	SOLTEIRO	SOLTEIRA
GRAU_INSTRUCAO	SUPERIOR COMPLETO	1º GRAU COMPLETO	2º GRAU COMPLETO	2º GRAU COMPLETO
CAT_PRETENDIDA	B	D	AB	AB
ATIV_REMUNERADA	N	S	N	N
NOME_CFC	CFC AB C	CFC AB P	CFC AB R	CFC AB V
MOT_REQUE	1º HAB	MUDANÇA DE CAT	1º HAB	1º HAB
RESULTADO	REPROVAD O	APROVADO	APROVADO	REPROVADO

Na Tabela 2 são apresentados 4 (quatro)*cluster*para os examinadores EX1520-EX1392 e observa-se os seguintes padrões de comportamento:

***Cluster*0**: o algoritmo identificou 24 (vinte e quatro) ocorrências semelhantes para este agrupamento, dentre as características estão, candidatos acima de 32 anos, do sexo feminino, solteira, com nível superior completo, categoria pretendida B, não exerce atividade remunerada, está matriculada na CFC AB C, primeira habilitação e que foram reprovados;

***Cluster*1**: o algoritmo identificou 15 (quinze) ocorrências semelhantes para este agrupamento, dentre as características estão, candidatos acima de 32 anos, sexo masculino, casados, com 1º grau completo, categoria pretendida D, exerce atividade remunerada, matriculados na CFC AB P, para mudança de categoria e foram aprovados no exame;

***Cluster*2**: o algoritmo identificou 61 (sessenta e uma) ocorrências semelhantes para este agrupamento, dentre as características estão, candidatos com idade entre 18 e 25 anos, sexo masculino, solteiros, com 2º grau completo, categoria pretendida AB, não exercem atividade remunerada, matriculados na CFC AB R, primeira habilitação e foram aprovados no exame;

***Cluster*3**: o algoritmo identificou 20 (vinte) ocorrências semelhantes para este agrupamento, dentre as características estão, candidatos acima de 32 anos, sexo feminino, solteiras, com 2º grau completo, categoria pretendida AB, não exercem atividade remunerada, matriculados na CFC AB V, primeira habilitação e foram reprovadas no exame;

Na Tabela 3 são apresentados os padrões obtidos ao aplicar o algoritmo de *cluster,*configurado para executar com 3 *cluster,* para os examinadores EX1011-EX0465 identificados como *outlier.*

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

Tabela 3. Tabela com resultado do cluster para o examinador EX1011-EX0465.

CAMPOS	CLUSTER 0(87)	CLUSTER 1 (26)	CLUSTER 2 (29)
IDADE	>32	>32	18-25
SEXO	FEMININO	MASCULINO	MASCULINO
ESTADO_CIVIL	CASADA	SOLTEIRO	SOLTEIRO
GRAU_INSTRUCAO	2º GRAU COMPLETO	1º GRAU INCOMPLETO	2º GRAU COMPLETO
CATEGORIA_PRETENDIDA	A	AB	AB
ATIVIDADEREMUNERADA	N	N	N
NOME_CFC	CFC AB T	CFC AB T	CFC AB T
MOTIVO_REQUERIMENTO	1º HAB	1º HAB	1º HAB
RESULTADO	APROVADO	APROVADO	APROVADO

É possível observar os seguintes padrões na Tabela 3:

***Cluster* 0** : o algoritmo identificou 87 (oitenta e sete) ocorrências semelhantes para este agrupamento, dentre as características estão, candidatos acima de 32 anos, do sexo feminino, casadas, com 2º grau completo, categoria pretendida A, não exerce atividade remunerada, está matriculada na CFC AB T, primeira habilitação e que foram aprovados;

***Cluster* 1** : o algoritmo identificou 26 (vinte e seis) ocorrências semelhantes para este agrupamento, dentre as características estão, candidatos acima de 32 anos, sexo masculino, solteiros, com 1º grau completo, categoria pretendida AB, não exerce atividade remunerada, matriculados na CFC AB T, para primeira habilitação e foram aprovados no exame;

***Cluster* 2** : o algoritmo identificou 29 (vinte e nove) ocorrências semelhantes para este agrupamento, dentre as características estão, candidatos com idade entre 18 e 25 anos, sexo masculino, solteiros, com 2º grau completo, categoria pretendida AB, não exercem atividade remunerada, matriculados na CFC AB T, primeira habilitação e foram aprovados no exame;

A equipe para trabalhar com o processo de extração de informações em base de dados é multidisciplinar, ou seja, há especialistas em ferramentas, recursos tecnológicos e de negócio.

Após extrair as informações e padrões da base de dados, elas foram repassadas para uma equipe de negócio formada por um analista com mais de 20 anos de experiência no setor de habilitação do Detran e o autor deste trabalho. Essa equipe interpretou e validou as informações, se são válidas ou não, e se as informações são úteis para atingir o objetivo deste trabalho.

Após análise dessas informações extraídas, conclui-se as seguintes situações:

1ª Interpretação : Após uma análise dos padrões extraídos na Ttabela 2, a equipe concluiu que não há uma relação entre examinadores e CFC, ou com gênero do candidato(a), categoria pretendida, atividade remunerada ou com qualquer outro atributo.

2ª Interpretação : Na Tabela 3, foi observado uma forte relação entre a dupla de examinadores e o CFC AB T, assim, identificando um forte indício de suspeita nos exames de direção. Sugere-se para este caso seja encaminhado para o DETRAN para analisar o caso e talvez sugerir uma auditoria detalhada do CFC e examinadores.

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

6. Conclusão

A mineração de dados é uma área que vem sendo disseminada cada vez mais por conta do crescente número de dados gerados e coletados pelas empresas e organizações. Além destes motivos para o crescimento dessa área, o aumento da capacidade computacional com baixo custo, métodos de aprendizado de máquina, análise e estatística de dados, possibilitam que as empresas tenham autonomia sobre o que fazer com os dados armazenados, assim, montando estratégias de como usar estes dados a seu favor.

Nos DETRANs existem inúmeros recursos para tentar coibir as fraudes para obter a CNH, no entanto, os criminosos, podendo ser as empresas, funcionários entre outros, elaboram novas formas para burlar o sistema. Além disso, os sistemas não são capazes de detectar atos humanos, por exemplo, o suborno. Esses atos são identificados muitas vezes por denúncias, o que acarreta posteriormente uma intensa investigação para identificar todos os envolvidos, como funciona e, posteriormente ações para inibir esses esquemas.

Diante deste cenário, os recursos de mineração de dados podem auxiliar a identificar essas fraudes, mesmo sendo praticadas de forma manual, ou seja, por atos do homem, sem a necessidade de burlar o sistema de informação dos DETRANs. Com os resultados obtidos foi possível identificar: examinadores que possuem índice de aprovação diferente dos padrões; identificar algumas relações entre esses examinadores e CFC, indicando que as autoridades devam analisar melhor a situação; e identificar CFC que já estavam envolvidas em problemas, de acordo com análise do especialista na área de negócio.

Estes resultados também podem auxiliar as autoridades competentes para instauração de auditoria, para se confirmar que as suspeitas são verídicas.

Algumas das limitações na execução deste trabalho foi a falta de uma base de dados com casos identificados como fraudes e a participação de órgãos competentes poderia acrescentar nas análises dos resultados.

Como trabalho futuro, é possível aplicar a metodologia deste trabalho em outras etapas do processo da CNH.

ISSN – 2317 - 0115
v. 6 , n. 2. (2017)
Recife – PE – UFPE / CCSA - MGP

Referências

Amo, Sandra de. (2004) "Técnicas de Mineração de Dados". UFU. In: XXIV Congresso da Sociedade Brasileira de Computação. Novembro.

BRASIL. Lei n. 9.503, de 23 de Setembro de 1997. Brasília, (1997), http://www.planalto.gov.br/ccivil_03/LEIS/L9503.htm , Novembro 2015.

Damasceno, Marcelo. (2010) "Introdução a mineração de dados utilizando o weka" Congresso de Pesquisa e Inovação da Rede Norte Nordeste de Educação Tecnológica. IFRN.

Fayyad, Usama et al. (1996) "The KDD Process for Extracting Useful Knowledge from Volumes of Data", In: Communications of the ACM. New York, v. 39, n. 11, p.27-39.

Gomes, Gabriel; Zoby, Leticia Toledo M. (2016). "Identificação de Padrões de Fraudes Na Obtenção Da Cnh, Utilizando Mineração", Trabalho de Conclusão de Curso, Centro Universitário IESB.

Han, J.; Kamber, M; Pei, J. (2012) "Data Mining Concepts and Techniques." USA: Elsevier. p.703

Hekima. (2014) "Por que a mineração de dados é essencial para as empresas que querem se destacar?", http://bigdatabusiness.com.br/porque-a-mineracao-de-dados-e-essencial-para-as-empresasque-querem-sedestacar , Agosto.

Maimon, Oded; Rokach, Lior. (2010) "Data Mining and Knowledge Discovery Handbook." New York: Springer.

Portal Brasil. (2014) "Carteira Nacional de Habilitação (CNH) possui cinco categorias", http://www.brasil.gov.br/cidadania-e-justica/2009/10/carteiranacional-de-habilitacao-cnh-possui-cinco-categorias , Novembro.

RSTUDIO. (2016) "Take control of your R code", https://www.rstudio.com/products/RStudio/ , Maio.

Stacciarine, Isa; Saraiva, Jacqueline; Cardim, Maria E. (2016) "Quadrilha cobrava até R$ 6 mil para fraudar exames do Detran-DF", http://www.correiobraziliense.com.br/app/noticia/cidades/2016/02/24/interna_cidadesdf,519065/quadrilha-cobrava-ate-r-6-mil-para-fraudar-exames-do-detran-df.shtml , Fevereiro.