Detecção de Fraudes – Uma Revisão Sistemática

Aluno: Jean Avila Rangel

Orientadores: Adolfo Neto e Maria Claudia Emer

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ (UTFPR)

Seminário 1

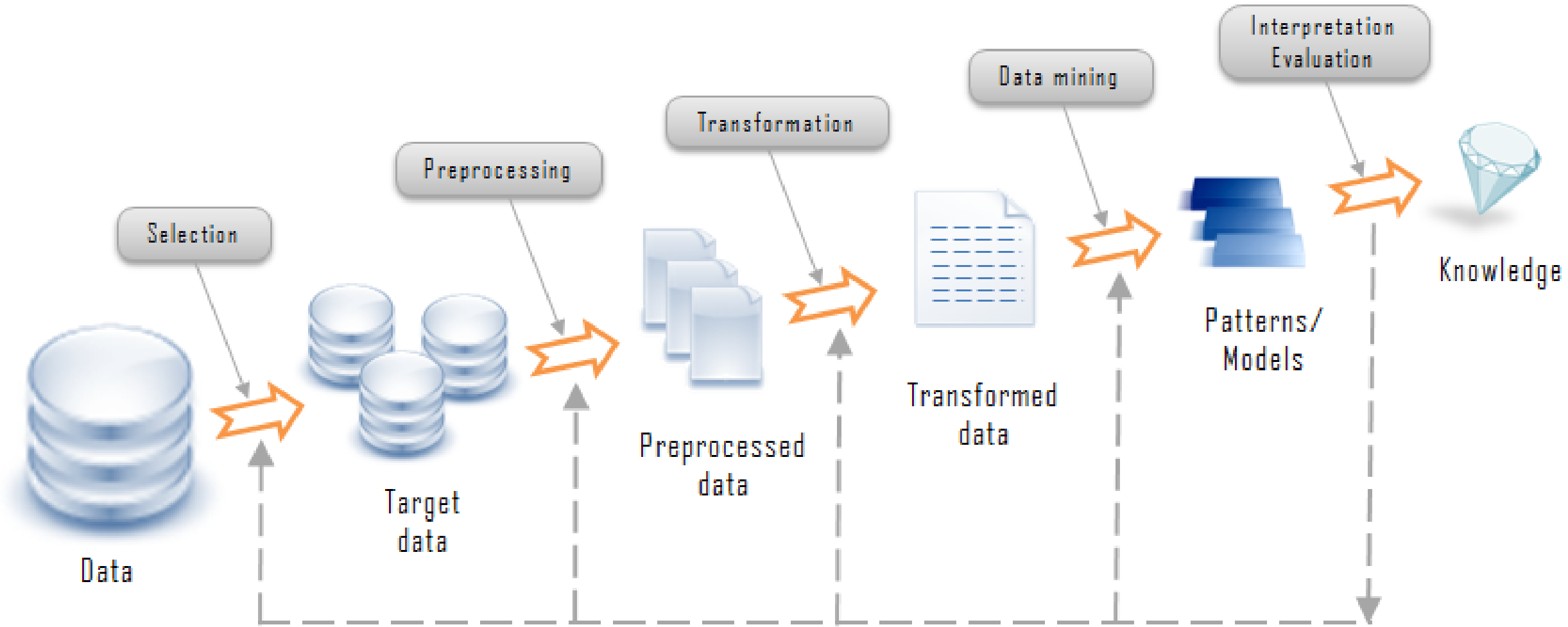
Curitiba, 2016

Agenda

- Tema;
- Motivação e justificativa;
- Objetivos;
  - Geral;
  - Específicos.
- Metodologia;
- Referencial teórico e estado da arte;
  - Áreas de detecção de fraudes;
  - Técnicas de detecção de fraudes.
- Conclusão.

Tema

- Uma atividade fraudulenta se caracteriza pela ação de um ou mais pessoas para ganhar proveito individual de determinada situação [1];
- A detecção de fraudes é uma área que visa encontrar falhas ou comportamentos inesperados em dados [4];
- O estudo da detecção de fraude pode englobar muitas áreas, como cartões de crédito e sistemas médicos;
- A um método computacional muito utilizado para detectar fraudes é a **mineração de dados**.


Selection
Preprocessing
Transformation
Data mining
Interpretation
Evaluation
Data
Target
data
Preprocessed
data
Transformed
data
Patterns/
Models
Knowledge

Motivação e justificativa

- Estudar uma área que está em constante acréscimo de importância no estado da arte;
- Há um grande número de revisões sistemáticas na área;
- Em trabalhos futuros, pretende-se aplicar ferramentas para detectar fraudes em dados de auditoria e controle de órgãos públicos.

Gráfico de publicações na ACM

Resultados em bases com as palavras-chave "Fraud Detection Survey" no periodo entre 2006 e 2016:

- **ACM Digital Library – 8 resultados;**
- **IEEEXplore – 51 resultados;**
- **ScienceDirect – 14 resultados;**
- **Springer Link – 170 resultados;**
- Google Scholar – 17.600 resultados.

Objetivos

- Geral;
  - O objetivo geral deste trabalho é realizar uma revisão sistemática para identificar e categorizar técnicas e ferramentas para detecção de fraudes dentro de áreas distintas.

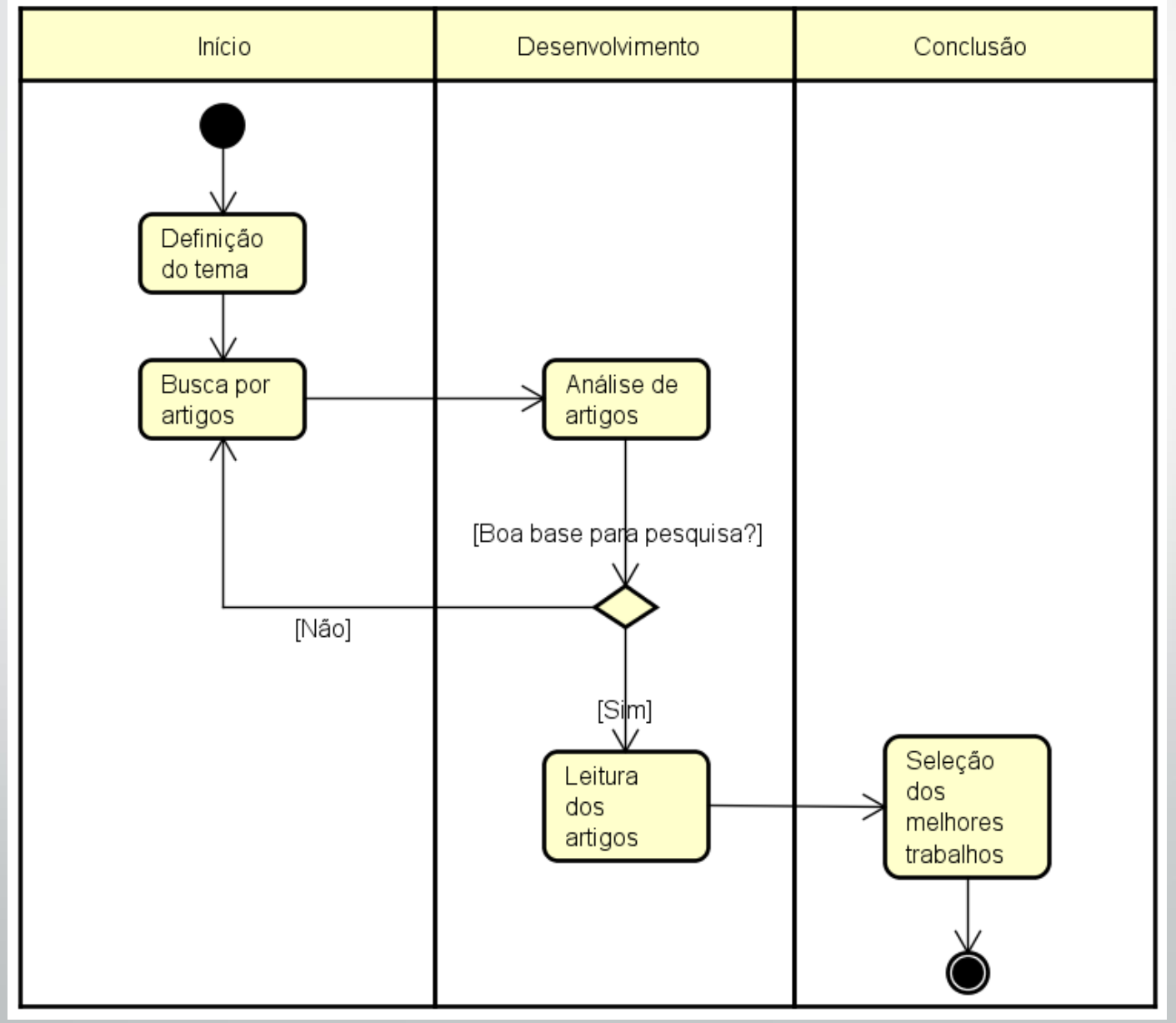
Objetivos

- Específicos.
  - Descobrir a importância das áreas de mineração de dados e aprendizado de máquina e suas utilizações em detecções de fraudes
  - Examinar se os estudos realizados em detecção de fraudes possuem abertura para áreas pouco ou não abordadas. Caso haja necessidade, o trabalho poderá categorizar possíveis trabalhos futuros em lacunas no assunto;
  - Elencar como os autores validaram suas pesquisas nas áreas de detecção de fraudes;
  - Categorizar quais tecnologias foram as mais utilizadas para cada contexto.

Metodologia

- Realizar um levantamento no estado da arte buscando aproximar do objetivo encontrado [7];
- Desenvolver a procura em bases de dados globais na área da informática científica.

Método de Pesquisa

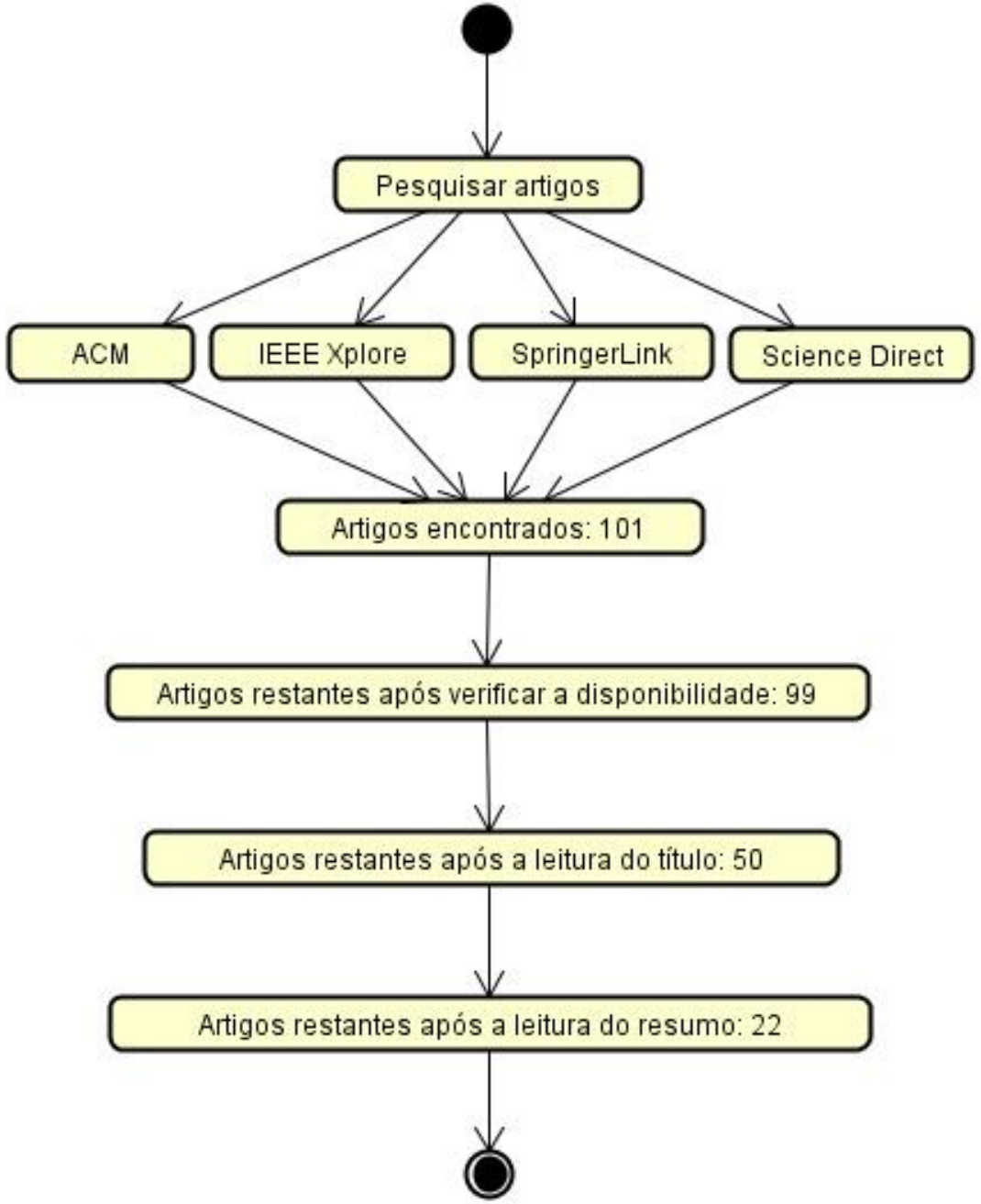


Critérios para exclusão de artigos

- Foram considerados somente trabalho com relevância científica, na língua inglesa e no período de 2006 a 2016;
- Após a leitura do título, resumo e artigo completo, foram excluídos os artigos que possuíam como tema central:
  - Imagens, vídeos ou impressões físicas;
  - Contextos biológicos, como análises de DNA e moléculas;
  - Componentes eletrônicos;
  - Dispositivos de segurança eletrônica relacionados a vírus, firewalls e ataques cibernéticos.

Foram desconsideradas as utilizações das palavras chave

- fraud detection **AND** state of art **OR** systematic review **OR** meta analysis;
- anomaly detection **AND** survey **OR** state of art **OR** systematic review **OR** meta analysis;
- outlier detection **AND** survey **OR** state of art **OR** systematic review **OR** meta analysis;
- deception detection **AND** survey **OR** state of art **OR** systematic review **OR** meta analysis.



Categoria	Quantia	Porcentagem
Conferência	12	55%
Revista	10	45%

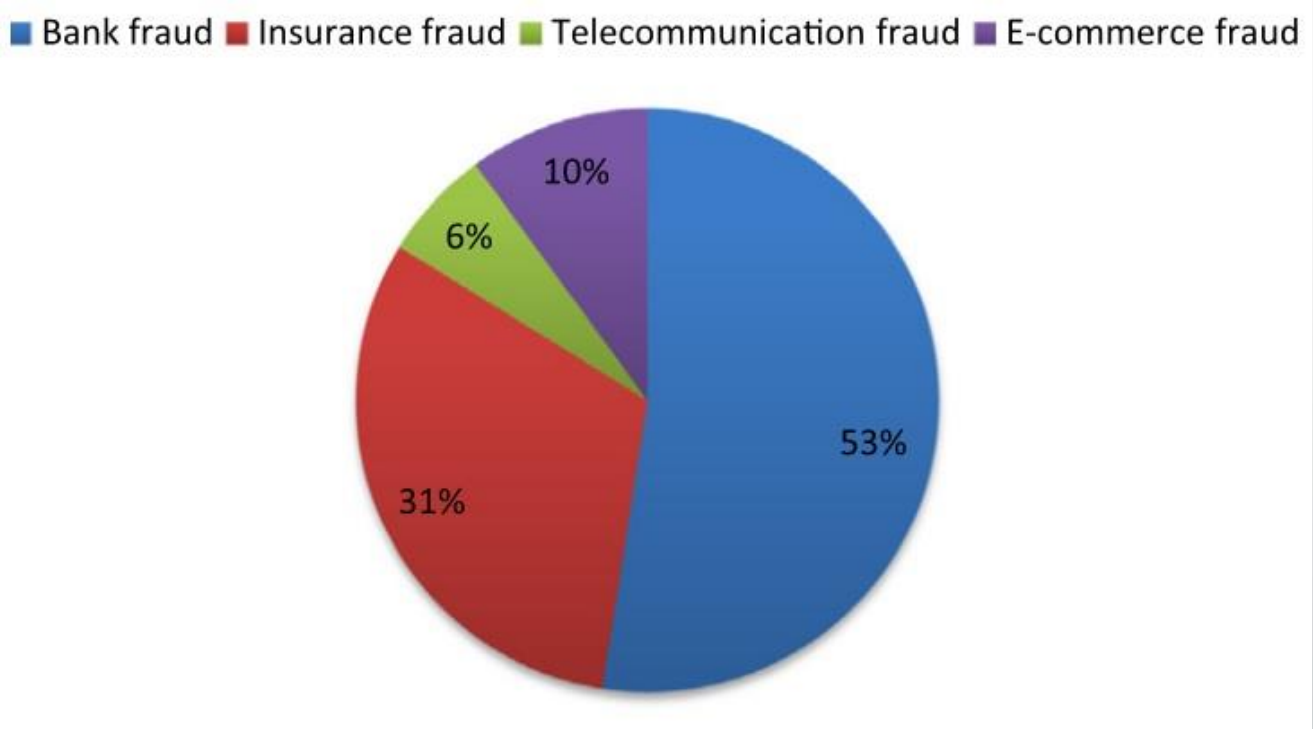
Questões de pesquisa

- Quais são as áreas de detecção de fraudes mais estudadas na literatura?
- Como a literatura categoriza as técnicas de detecção de fraudes?
- Quais foram os problemas mais relatados pelos autores?
- Como os autores testaram e validaram suas pesquisas?
- Em qual área há pouca abordagem no estudo de detecção de fraudes?
- Há espaço para futuras pesquisas na área?

Agenda

- Tema;
- Motivação e justificativa;
- Objetivos;
  - Geral;
  - Específicos.
- Metodologia;
- **Referencial teórico e estado da arte;**
  - **Áreas de detecção de fraudes;**
  - **Técnicas para detecção de fraudes.**
- Conclusão.

Áreas mais abordadas em detecção de fraudes



- Fonte: Abdallah et al (2016)

Áreas mais abordadas em detecção de fraudes

- Pejic-Bach M (2010) [9] analisou artigos sobre detecção de fraudes desde 1956 e notou a maior concentração em sistemas financeiros, sistemas de telecomunicação e seguradoras;

Áreas	Autores
Cartão de crédito	Abdallah et al [2016], Bansal et al [2016], Ahmed et al [2015], Edge and Falcone Sampaio [2009], Allan and Zhan [2010], Pejic-Bach [2010], Raj and Portia [2011], Perlich et al [2007], Kanapickienė et al [2015], Branco [2016], Akoglu et al [2015]
Seguradoras	Abdallah et al [2016], Bansal et al [2016], Ahmed et al [2015], Pejic-Bach [2010], Branco [2016], Akoglu et al [2015]
Cuidados da saúde	Abdallah et al [2016], Bauder et al [2016], Li et al [2008], Bansal et al [2016], Pejic-Bach [2010]
Telecomunicação	Abdallah et al [2016], Ahmed et al [2015], Allan and Zhan [2010], Pejic-Bach [2010], Pejic-Bach [2010], Branco [2016], Akoglu et al [2015]
Comércio eletrônico	Abdallah et al [2016], Pejic-Bach [2010], Branco [2016], Akoglu et al [2015]
Leilões virtuais	Abdallah et al [2016]
Mercados de ações	Ahmed et al [2015]
Dados financeiros ou empresariais	Flegel et al [2010], Edge and Falcone Sampaio [2009], Gullkvist and Jokipii [2013], Allan and Zhan [2010], Pejic-Bach [2010], Wang [2010], Branco [2016], Akoglu et al [2015]
Redes sociais	Yu et al [2016], Liu and Chawla [2015], Feily et al [2009], Rebahi et al [2011]
Big data	Sharma and Mangat [2015], Mahmood and Afzal [2013]
Redes de computadores	Allan and Zhan [2010], Branco [2016], Akoglu et al [2015]
Equipamentos industriais	Bauder et al [2016]
Terrorismo	Allan and Zhan [2010]

Objetivos da utilização de sistemas para detectar fraudes

- Redução de custos com usuários maliciosos [11];
- Alertas de dados errôneos ou detentores de ruídos;
- Visualização de possíveis padrões indesejados;
- Melhor monitoramento das informações.

Técnicas para detecção de fraudes mais utilizadas

- Método supervisionado baseado em aprendizado (classificação):
  - Ex: redes neurais;
- Método não supervisionado (clusterização ou classificação):
  - Ex: *K-Means*;

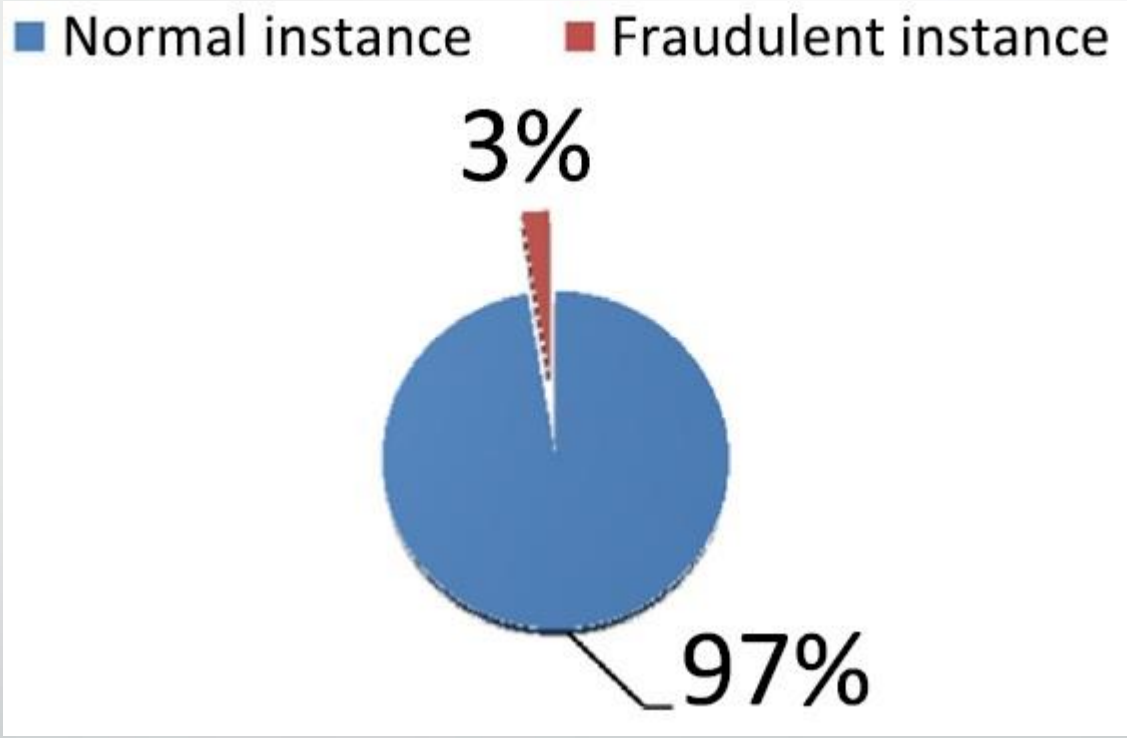
Tipos de anomalias

- Indivíduo fora do padrão;
- Indivíduo fora do padrão por contexto;
- Grupo fora do padrão.

Categorias e técnicas utilizadas em métodos supervisionados e não-supervisionados

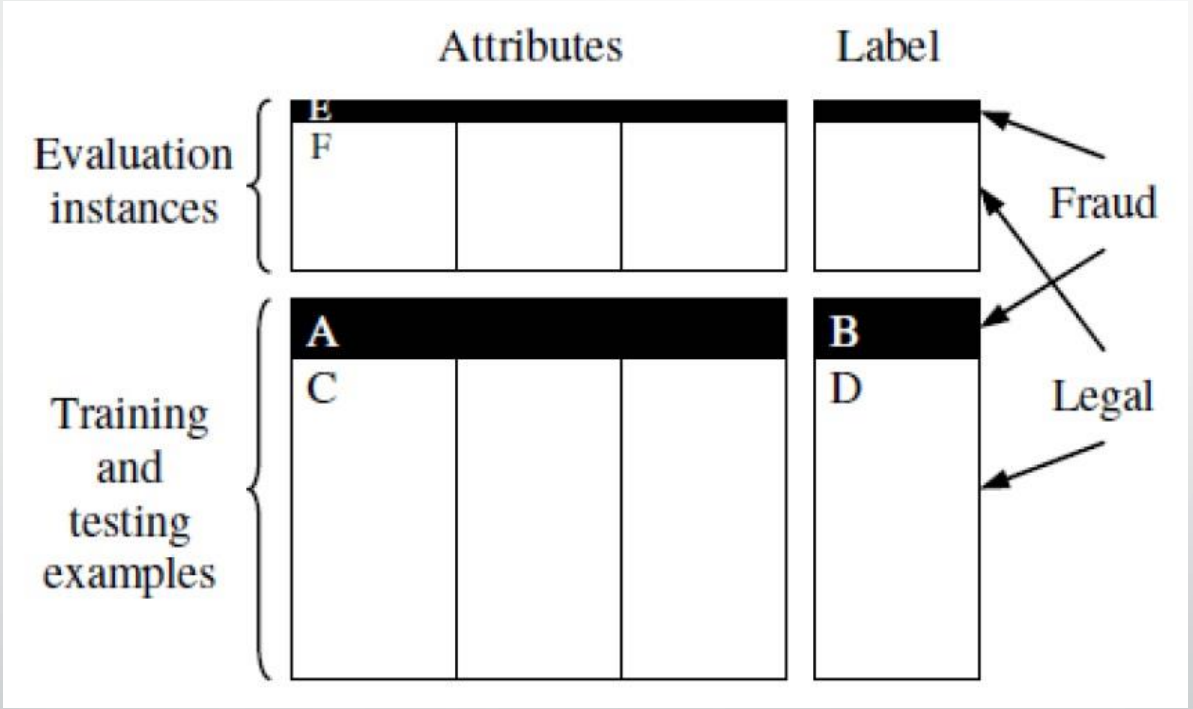
Método	Categoria	Técnica
Supervisionado	Classificação	Algoritmo genético Árvores de decisões Baseado em regras Inferência Bayesiana K-Nearest neighbours (K-NN) Naive Bayes Modelo oculto de Markov Redes neurais artificiais Support vector machine (SVM)
Não-supervisionado	Clusterização	Baseado em distância K-Means Lógica Fuzzy Mistura gaussiana Principal component analysis (PCA)

Quantia de dados fraudulentos em uma base real



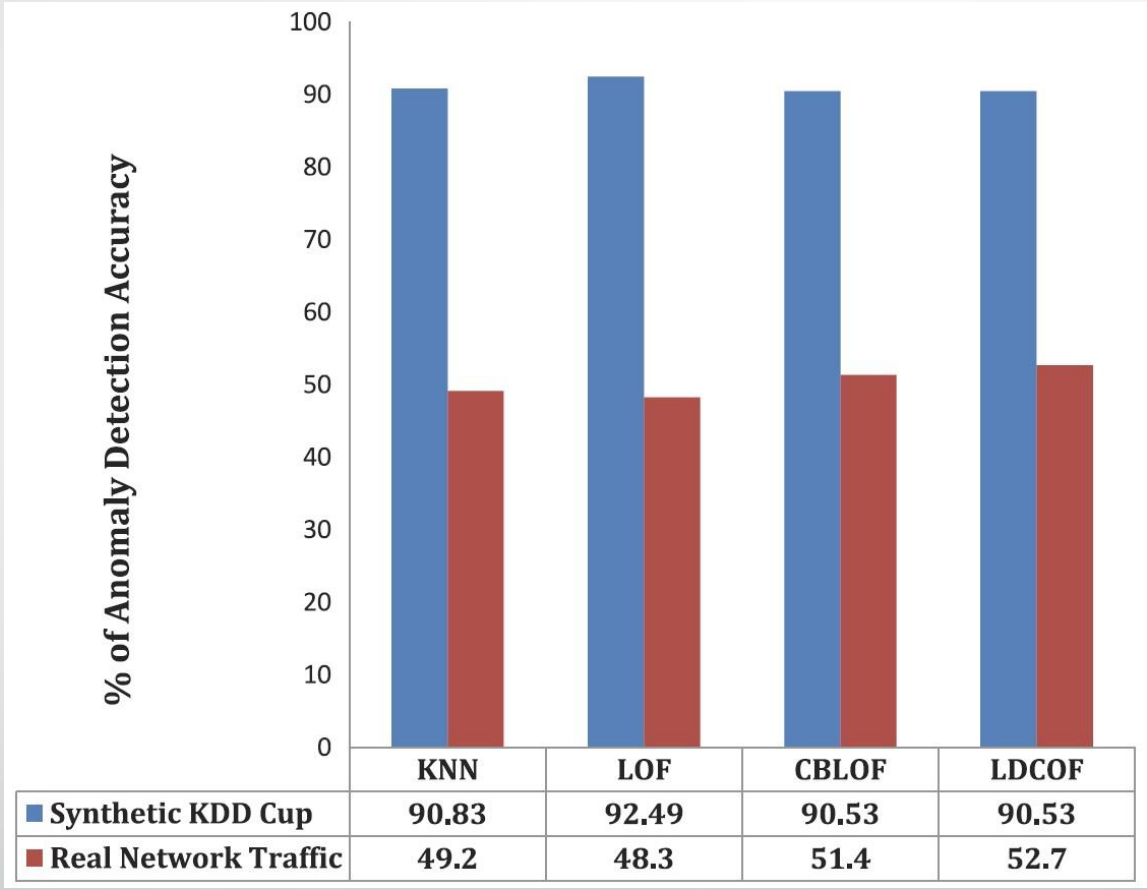
- Fonte: Abdallah et al (2016)

Conjunto de treinamento para métodos supervisionados



- Fonte: Allan e Zhan (2010)

Dados reais e dados fictícios



	KNN	LOF	CBLOF	LDCOF
Synthetic KDD Cup	90.83	92.49	90.53	90.53
Real Network Traffic	49.2	48.3	51.4	52.7

- Fonte: Ahmed et al (2015)

Conclusão: respostas para as questões de pesquisa

- Quais são as áreas de detecção de fraudes mais estudadas na literatura?
- Como a literatura categoriza as técnicas de detecção de fraudes?
- Quais foram os problemas mais relatados pelos autores?
- Como os autores testaram e validaram suas pesquisas?
- Em qual área há pouca abordagem no estudo de detecção de fraudes?
- Há espaço para futuras pesquisas na área?

Operação Serenata de Amor - https://serenata.datasciencebr.com/

Operação Serenata de Amor

ROBÔS LUTANDO CONTRA A CORRUPÇÃO NO BRASIL

A Operação Serenata de Amor nasceu de uma combinação de necessidades de muitas pessoas: de ver Aprendizado de Máquina aplicado para melhorar nossas vidas, de aprender em quem devemos votar e de fazer algo a respeito do problema de corrupção que afeta o mundo inteiro.

Estamos construindo uma inteligência artificial capaz de analisar contas públicas e de dizer, com confiança, a possibilidade de cada nota ser ilegal. Tudo que estamos construindo está sendo feito com código aberto desde o início, onde você, como tantos outros já fazem, pode contribuir ativamente com a construção do projeto.

Operação Serenata de Amor

Nosso objetivo no momento é conseguir aplicar essa inteligência na Cota para Exercício da Atividade Parlamentar (CEAP), da câmara dos deputados no Brasil. Esse trabalho involve a criação de APIs, limpeza de dados e análises, concepção e validação de hipóteses e confirmação de gastos ilícitos através de investigações.

Para chegar nesse ponto, inédito, convidamos a todos para treinar a inteligência artificial, coletar informações, cruzar bancos de dados, validar hipóteses e aplicar Aprendizado de Máquina com modelos competindo entre si - todos buscando ter uma precisão maior do que o anterior.

Referências

- Abdallah A, Maarof M, Zainal A (2016) Fraud detection system: A survey. Journal of Network and Computer Applications 68:90 – 113;
- Ahmed M, Mahmood A, Islam M (2015) A survey of anomaly detection techniques in financial domain. Future Generation Computer Systems, 55:278–288.
- Allan T, Zhan J (2010) Towards fraud detection methodologies. 2010 5th International Conference on Future Information Technology, FutureTech 2010 - Proceedings.
- Chan PK, Fan W, Prodromidis AL, Stolfo SJ (1999) Distributed data mining in credit card fraud detection. IEEE Intelligent Systems and their Applications 14(6):67–74;
- Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. ACM Comput Surv 41(3):15:1–15:58;
- Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI magazine, v. 17, n. 3, p. 37;
- Fawcett T, Provost F (1997) Adaptive fraud detection. Data Mining and Knowledge Discovery 1(3):291–316;
- Kitchenham B (2004) Procedures for performing systematic reviews. Keele, UK, Keele University;
- Pejic-Bach M (2010) Invited Paper: Profiling Intelligent Systems Applications in Fraud Detection and Prevention: Survey of Research Articles, *2010 International Conference on Intelligent Systems, Modelling and Simulation*, Liverpool, pp. 80-85;

  Seyedhossein L, Hashemi MR (2010) Mining information from credit card time series for timelier fraud detection. In: Telecommunications (IST), 2010 5th International Symposium on, pp 619–624.