

# Detecção inteligente de efeitos colaterais indesejáveis na Internet das coisas - um estudo de caso no Home Network System

## Apresentação de Monografia

Heron Sanches Gonçalves Pires Ferreira



Universidade Federal da Bahia - Departamento de Ciência da Computação

**Orientadora:** Profa. Dra. Daniela Barreiro Claro

**Co-Orientador:** Roberto Cerqueira Figueiredo

Contato: heronsanches@dcc.ufba.br

31 de outubro de 2016

# Conteúdo

## 1 Introdução

## 2 Proposta

- Cenário Levar compras
- Método de detecção
  - Geração da massa de dados
  - Seleção dos atributos
  - Método ensemble DECORATE
- Implantação do método proposto

## 3 Validação

- Método avaliativo stratified-k-fold-cross-validation
- Experimentos
- Resultados e discussão
  - Modelos
  - Implantação

## 4 Conclusão e trabalhos futuros

# Conteúdo

## 1 Introdução

## 2 Proposta

- Cenário Levar compras
- Método de detecção
  - Geração da massa de dados
  - Seleção dos atributos
  - Método ensemble DECORATE
- Implantação do método proposto

## 3 Validação

- Método avaliativo stratified-k-fold-cross-validation
- Experimentos
- Resultados e discussão
  - Modelos
  - Implantação

## 4 Conclusão e trabalhos futuros

# Introdução

- A Internet está se tornando cada vez mais persistente no cotidiano (Chandrakanth et al., 2014).
- Em 2010 havia aproximadamente 1,5 bilhão de PCs conectados a Internet e mais que 1 bilhão de telefones móveis (Sundmaeker et al., 2010).
- Segundo Gartner<sup>1</sup>, 6,4 bilhões de coisas estarão conectadas até o final de 2016 e, em 2020 esse número atingirá cerca de 20,8 bilhões.

---

<sup>1</sup><<http://www.gartner.com/newsroom/id/3165317>>

# Introdução

- A Internet está se tornando cada vez mais persistente no cotidiano (Chandrakanth et al., 2014).
- Em 2010 havia aproximadamente 1,5 bilhão de PCs conectados a Internet e mais que 1 bilhão de telefones móveis (Sundmaeker et al., 2010).
- Segundo Gartner<sup>1</sup>, 6,4 bilhões de coisas estarão conectadas até o final de 2016 e, em 2020 esse número atingirá cerca de 20,8 bilhões.

---

<sup>1</sup><<http://www.gartner.com/newsroom/id/3165317>>

# Introdução

- A Internet está se tornando cada vez mais persistente no cotidiano (Chandrakanth et al., 2014).
- Em 2010 havia aproximadamente 1,5 bilhão de PCs conectados a Internet e mais que 1 bilhão de telefones móveis (Sundmaeker et al., 2010).
- Segundo Gartner<sup>1</sup>, 6,4 bilhões de coisas estarão conectadas até o final de 2016 e, em 2020 esse número atingirá cerca de 20,8 bilhões.

---

<sup>1</sup><<http://www.gartner.com/newsroom/id/3165317>>

# Coisa: veículo, animal, dispositivo doméstico, pessoa, ...

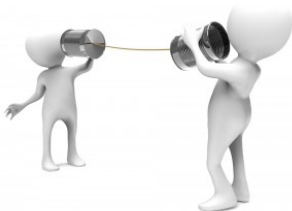


Imagem: Baseada em (Barrett, 2012)

# Internet das Coisas (IoT)

Rede mundial de objetos (coisas) unicamente endereçáveis e interconectados, seguindo os protocolos dos padrões de comunicação (Enterprise et al., 2008).

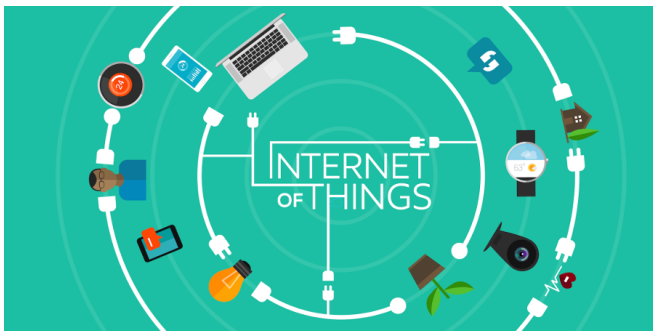


Imagem: <sup>2</sup>

---

<sup>2</sup><<http://intca.org/2016/08/internet-of-things/>>



# Internet das Coisas (IoT)

Devido ao crescente número de coisas sendo conectadas, surgem diversos desafios, a exemplo de:

- Disponibilidade de uma interface de comunicação (acesso aos serviços e informações dos dispositivos) comum aos objetos.
- Detecção e resolução de efeitos colaterais indesejáveis.

# Internet das Coisas (IoT)

Devido ao crescente número de coisas sendo conectadas, surgem diversos desafios, a exemplo de:

- Disponibilidade de uma interface de comunicação (acesso aos serviços e informações dos dispositivos) comum aos objetos.
- Detecção e resolução de efeitos colaterais indesejáveis.

# Desafios IoT: interface de comunicação padrão

Disponibilização das coisas como serviços Web RESTful.

- Mais leve que os serviços Web baseados em SOAP.

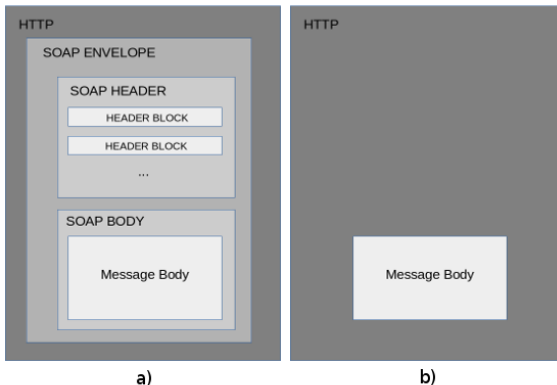


Imagem: Baseada em (Pautasso, 2014)

# Desafios IoT: interface de comunicação padrão

Disponibilização das coisas como serviços Web RESTFul.

- Diferentes representações de um mesmo recurso (aumenta a interoperabilidade)

## Exemplos

JSON, XML, TEXT, TEXT\_HTML.

- Interface com semântica bem definida.

## Exemplos

Utilização dos métodos HTTP (GET, PUT, DELETE, POST, HEAD, OPTIONS, dentre outros).

# Desafios IoT: interface de comunicação padrão

Disponibilização das coisas como serviços Web RESTFul.

- Diferentes representações de um mesmo recurso (aumenta a interoperabilidade)

## Exemplos

JSON, XML, TEXT, TEXT\_HTML.

- Interface com semântica bem definida.

## Exemplos

Utilização dos métodos HTTP (GET, PUT, DELETE, POST, HEAD, OPTIONS, dentre outros).

# Desafios IoT: detecção e resolução de efeitos colaterais indesejáveis

Em desenvolvimento de *software*, uma *feature* (característica) é um componente de adicional funcionalidade ao *software* (Calder et al., 2003), consistindo de um conjunto de requisitos logicamente relacionados e suas especificações, o qual se destina a fornecer um determinado efeito comportamental (NHLABATSI et al., 2008).

# Desafios IoT: detecção e resolução de efeitos colaterais indesejáveis

Quando a **composição de *features*** leva a algum **comportamento não esperado** - interação de características, esta pode resultar em **efeitos colaterais indesejáveis**: um estado inconsistente do sistema, um sistema instável ou dados imprecisos (NHLABATSI et al., 2008).

# Conteúdo

## 1 Introdução

## 2 Proposta

- Cenário Levar compras
- Método de detecção
  - Geração da massa de dados
  - Seleção dos atributos
  - Método ensemble DECORATE
- Implantação do método proposto

## 3 Validação

- Método avaliativo stratified-k-fold-cross-validation
- Experimentos
- Resultados e discussão
  - Modelos
  - Implantação

## 4 Conclusão e trabalhos futuros

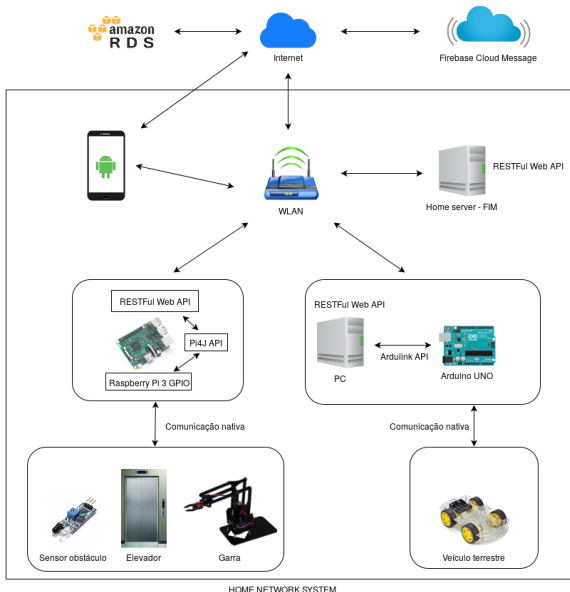


# Proposta

Detectar efeitos colaterais indesejáveis de uma maneira inteligente, utilizando o método *ensemble* DECORATE, no cenário “Levar compras” do *Home Network System* (HNS).

Rede doméstica de coisas (aparelhos domésticos e sensores) com capacidade de conectividade de rede e, interface de controle e monitoramento. Os dispositivos da casa são compostos uns com os outros para prover funcionalidades que atendam aos requisitos dos usuários da casa (Nakamura et al., 2013).

# HNS do Cenário Levar compras



# Objetos das compras (cestas)

| largura | altura | comprimento | massa | diâmetro | nome                             |
|---------|--------|-------------|-------|----------|----------------------------------|
| 0       | 0.031  | 0           | 0.003 | 0.009    | Refrigerante nº 4 - Soda         |
| 0.01    | 0.014  | 0.03        | 0.001 | 0        | Caixa Pasta de dente - vermelha. |
| 0.004   | 0.02   | 0.014       | 0.001 | 0        | Batata                           |
| 0.004   | 0.02   | 0.014       | 0.001 | 0        | Café                             |
| 0       | 0.015  | 0           | 0.003 | 0.008    | Café Classic                     |
| 0.005   | 0.02   | 0.011       | 0.001 | 0        | Cerveja em lata III              |
| 0       | 0.028  | 0           | 0.001 | 0.007    | Champagne - Azul                 |
| 0.004   | 0.02   | 0.014       | 0.001 | 0        | Farinha de Trigo                 |
| 0       | 0.03   | 0           | 0.001 | 0.008    | Garrafa de Rum                   |
| 0.012   | 0.016  | 0.004       | 0.004 | 0        | Goiabada                         |
| 0       | 0.03   | 0           | 0.001 | 0.008    | Garrafa de Vodka Special         |
| 0.027   | 0.014  | 0.018       | 0.001 | 0        | Pote de sorvete                  |
| 0       | 0.032  | 0           | 0.002 | 0.007    | Vinho branco                     |
| 0.012   | 0.03   | 0.006       | 0.002 | 0        | Limpa vidros                     |
| 0.021   | 0.015  | 0.007       | 0.002 | 0        | Sabonete - II                    |
| 0.015   | 0.024  | 0.008       | 0.001 | 0        | Sabão em pó - II                 |
| 0.019   | 0.03   | 0.039       | 0.018 | 0        | Bau simples                      |
| 0.046   | 0.061  | 0.017       | 0.006 | 0        | Prateleira de acrílico pequena   |
| 0.016   | 0.025  | 0.004       | 0.001 | 0        | Farinha de trigo                 |
| 0       | 0.03   | 0           | 0.003 | 0.022    | Barril em madeira                |
| 0.013   | 0.017  | 0.0063      | 0.002 | 0        | Fermento em pó                   |
| 0.077   | 0.0197 | 0.038       | 0.001 | 0        | Colchão p/ berço azul            |
| 0.01    | 0.0432 | 0           | 0.001 | 0        | Jogo de esfregões                |
| 0.0181  | 0.0258 | 0.0017      | 0.001 | 0        | Pão de Queijo                    |
| 0.022   | 0.0412 | 0.0135      | 0.001 | 0        | Rastelo                          |

**Imagem:** Objetos das compras (cestas). Medidas em metros (m).

# Exemplos de Compras (cestas)



Imagem: Exemplos de cestas.

# Cenário Levar compras

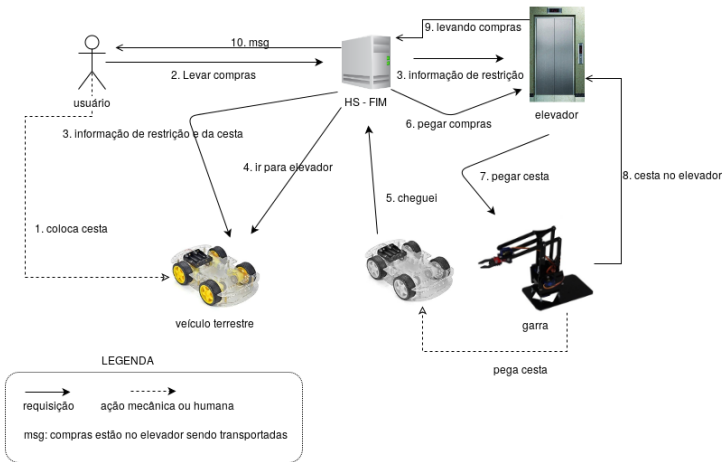


Imagem: Cenário Levar compras.

# Geração da massa de dados

- Confeccionadas 59 cestas.
- Verificado se as cestas cabiam dentro do elevador.
- Registro no banco de dados relacional na nuvem.
- Cenário Levar compras executado 59 vezes.

# Geração da massa de dados

- Confeccionadas 59 cestas.
- Verificado se as cestas cabiam dentro do elevador.
- Registro no banco de dados relacional na nuvem.
- Cenário Levar compras executado 59 vezes.



# Geração da massa de dados

- Confeccionadas 59 cestas.
- Verificado se as cestas cabiam dentro do elevador.
- Registro no banco de dados relacional na nuvem.
- Cenário Levar compras executado 59 vezes.

# Geração da massa de dados

- Confeccionadas 59 cestas.
- Verificado se as cestas cabiam dentro do elevador.
- Registro no banco de dados relacional na nuvem.
- Cenário Levar compras executado 59 vezes.

# Geração da massa de dados

- Algumas vezes o comportamento da garra fazia com que a cesta não ficasse corretamente no compartimento do elevador, impedindo o elevador continuar com a funcionalidade “Levar compras”. Provocando um **efeito colateral indesejável**.
- Cada cesta foi rotulada como “É efeito colateral” ou “não é efeito colateral” e atualizado valor no banco.

# Geração da massa de dados

- Algumas vezes o comportamento da garra fazia com que a cesta não ficasse corretamente no compartimento do elevador, impedindo o elevador continuar com a funcionalidade “Levar compras”. Provocando um **efeito colateral indesejável**.
- Cada cesta foi rotulada como “**É efeito colateral**” ou “**não é efeito colateral**” e atualizado valor no banco.

# Seleção dos atributos

- Uma cesta com 10 objetos tem  $10 \times 5 = 50$  atributos.
- Usado os valores das médias e totais das medidas, assim como a quantidade de objetos. Reduzindo a 11 atributos.

# Seleção dos atributos

- Uma cesta com 10 objetos tem  $10 \times 5 = 50$  atributos.
- Usado os valores das médias e totais da medidas, assim como a quantidade de objetos. Reduzindo a 11 atributos.

# Seleção dos atributos

- Construído arquivo “arff”, padrão WEKA (Hall et al., 2009), com 59 instâncias.
- Cada instância representa uma cesta com 11 parâmetros numéricos mais 1 nominal.
- Gerou-se uma visualização do espaço de atributos em 2D, ou seja, par a par.

# Seleção dos atributos

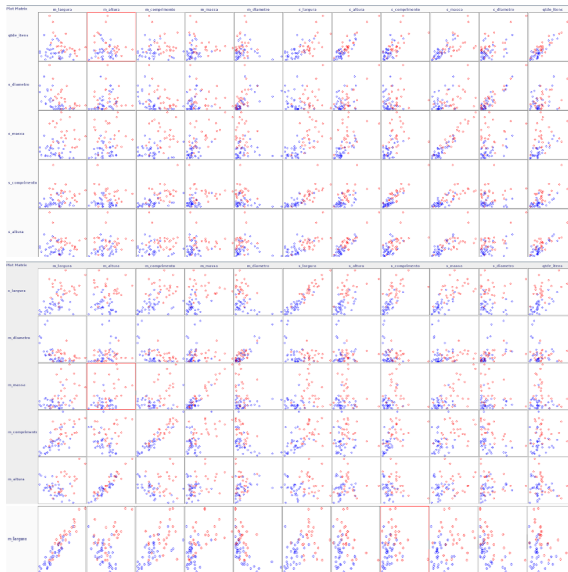
- Construído arquivo “arff”, padrão WEKA (Hall et al., 2009), com 59 instâncias.
- Cada instância representa uma cesta com 11 parâmetros numéricos mais 1 nominal.
- Gerou-se uma visualização do espaço de atributos em 2D, ou seja, par a par.



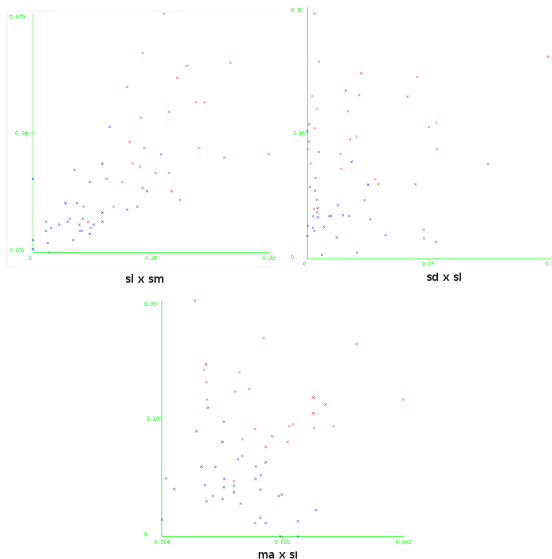
# Seleção dos atributos

- Construído arquivo “arff”, padrão WEKA (Hall et al., 2009), com 59 instâncias.
- Cada instância representa uma cesta com 11 parâmetros numéricos mais 1 nominal.
- Gerou-se uma visualização do espaço de atributos em 2D, ou seja, par a par.

# Seleção dos atributos



# Seleção dos atributos - 3 pares selecionados



# DECORATE (Melville e Mooney, 2003; Melville e Mooney, 2004)

DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples).

- Combina o resultado de diversos classificadores para chegar em uma decisão final.

$$P_y(x) = \frac{\sum_{C_i \in C^*} P_{C_{i,y}}(x)}{|C^*|} \quad (1)$$

# DECORATE

- Utiliza dados adicionais de treino gerados artificialmente.

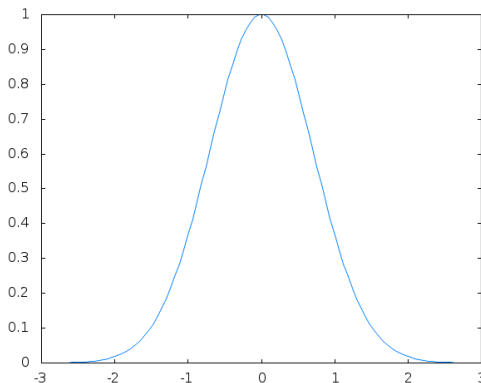


Imagem: Curva Gaussiana. <sup>3</sup>

---

<sup>3</sup><<http://proooof.blogspot.com.br/2011/12/sigma.html>>

# DECORATE

- Utiliza dados adicionais de treino gerados artificialmente.
- Primeiramente utiliza a probabilidade do *ensemble*  $P_y(x)$ .
- Por fim utiliza a expressão abaixo para rotular o exemplo artificial.

$$P'_y(x) = \frac{\frac{1}{P_y(x)}}{\sum_y \frac{1}{P_y(x)}} \quad (2)$$

# DECORATE

- Utiliza dados adicionais de treino gerados artificialmente.
- Primeiramente utiliza a probabilidade do *ensemble*  $P_y(x)$ .
- Por fim utiliza a expressão abaixo para rotular o exemplo artificial.

$$P'_y(x) = \frac{\frac{1}{P_y(x)}}{\sum_y \frac{1}{P_y(x)}} \quad (2)$$

# DECORATE

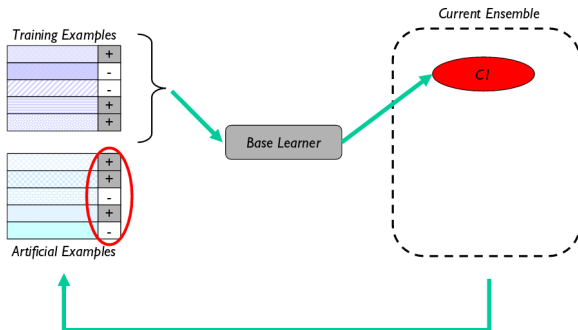


Imagem: DECORATE 1ª iteração. <sup>4</sup>

<sup>4</sup> <<http://www.time.mk/trajkovski/teaching/aim/Lecture8.pdf>>



# DECORATE

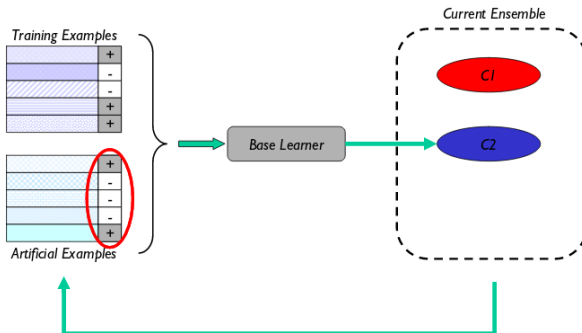


Imagem: DECORATE 2ª iteração. <sup>5</sup>

<sup>5</sup><<http://www.time.mk/trajkovski/teaching/aim/Lecture8.pdf>>

# DECORATE (Melville e Mooney, 2003; Melville e Mooney, 2004)

- Comparações (Firdausi et al., 2010; Arora e Suman, 2012; Karthikeyan e Thangaraju, 2013) entre J48 (Hall et al., 2009) e MLP (redes neurais artificiais) mostram que o J48 obteve resultados satisfatórios e mostrou-se ser muito menos custoso computacionalmente.
- DECORATE obteve melhores resultados quando comparado ao J48 atuando isoladamente.
- DECORATE tendo como base o J48 obteve taxas de *accuracy* satisfatórias quando comparado com outros *ensembles* e melhores em *datasets* pequenos.

# DECORATE (Melville e Mooney, 2003; Melville e Mooney, 2004)

- Comparações (Firdausi et al., 2010; Arora e Suman, 2012; Karthikeyan e Thangaraju, 2013) entre J48 (Hall et al., 2009) e MLP (redes neurais artificiais) mostram que o J48 obteve resultados satisfatórios e mostrou-se ser muito menos custoso computacionalmente.
- DECORATE obteve melhores resultados quando comparado ao J48 atuando isoladamente.
- DECORATE tendo como base o J48 obteve taxas de *accuracy* satisfatórias quando comparado com outros *ensembles* e melhores em *datasets* pequenos.

# DECORATE (Melville e Mooney, 2003; Melville e Mooney, 2004)

- Comparações (Firdausi et al., 2010; Arora e Suman, 2012; Karthikeyan e Thangaraju, 2013) entre J48 (Hall et al., 2009) e MLP (redes neurais artificiais) mostram que o J48 obteve resultados satisfatórios e mostrou-se ser muito menos custoso computacionalmente.
- DECORATE obteve melhores resultados quando comparado ao J48 atuando isoladamente.
- DECORATE tendo como base o J48 obteve taxas de *accuracy* satisfatórias quando comparado com outros *ensembles* e melhores em *datasets* pequenos.

# Parâmetros DECORATE (Melville e Mooney, 2004)

- Número máximo de iterações = 50.
- Número máximo de classificadores = 14.
- Classificador base = J48.
- Quantidade de exemplos artificiais = 100% do *dataset* de treino.

# Parâmetros DECORATE (Melville e Mooney, 2004)

- Número máximo de iterações = 50.
- Número máximo de classificadores = 14.
- Classificador base = J48.
- Quantidade de exemplos artificiais = 100% do *dataset* de treino.

# Parâmetros DECORATE (Melville e Mooney, 2004)

- Número máximo de iterações = 50.
- Número máximo de classificadores = 14.
- Classificador base = J48.
- Quantidade de exemplos artificiais = 100% do *dataset* de treino.

# Parâmetros DECORATE (Melville e Mooney, 2004)

- Número máximo de iterações = 50.
- Número máximo de classificadores = 14.
- Classificador base = J48.
- Quantidade de exemplos artificiais = 100% do *dataset* de treino.



# Implantação do método proposto

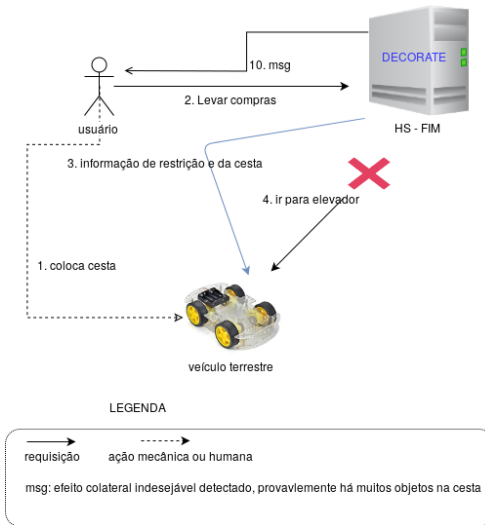


Imagem: Atuação do DECORATE no cenário Levar compras.

# Conteúdo

## 1 Introdução

## 2 Proposta

- Cenário Levar compras
- Método de detecção
  - Geração da massa de dados
  - Seleção dos atributos
  - Método ensemble DECORATE
- Implantação do método proposto

## 3 Validação

- Método avaliativo stratified-k-fold-cross-validation
- Experimentos
- Resultados e discussão
  - Modelos
  - Implantação

## 4 Conclusão e trabalhos futuros

# TP, TN, FP, FN (Olson e Delen, 2008)

- Verdadeiro positivo (TP): instância que pertence a classe positiva e que foi corretamente classificada como positiva.
- Verdadeiro negativo (TN): instância que pertence a classe negativa e que foi classificada corretamente como negativa.
- Falso positivo (FP): instância que pertence a classe negativa e que foi classificada incorretamente como positiva.
- Falso negativo (FN): instância que pertence a classe positiva e que foi classificada incorretamente como negativa.

# TP, TN, FP, FN (Olson e Delen, 2008)

- Verdadeiro positivo (TP): instância que pertence a classe positiva e que foi corretamente classificada como positiva.
- Verdadeiro negativo (TN): instância que pertence a classe negativa e que foi classificada corretamente como negativa.
- Falso positivo (FP): instância que pertence a classe negativa e que foi classificada incorretamente como positiva.
- Falso negativo (FN): instância que pertence a classe positiva e que foi classificada incorretamente como negativa.

# TP, TN, FP, FN (Olson e Delen, 2008)

- Verdadeiro positivo (TP): instância que pertence a classe positiva e que foi corretamente classificada como positiva.
- Verdadeiro negativo (TN): instância que pertence a classe negativa e que foi classificada corretamente como negativa.
- Falso positivo (FP): instância que pertence a classe negativa e que foi classificada incorretamente como positiva.
- Falso negativo (FN): instância que pertence a classe positiva e que foi classificada incorretamente como negativa.

# TP, TN, FP, FN (Olson e Delen, 2008)

- Verdadeiro positivo (TP): instância que pertence a classe positiva e que foi corretamente classificada como positiva.
- Verdadeiro negativo (TN): instância que pertence a classe negativa e que foi classificada corretamente como negativa.
- Falso positivo (FP): instância que pertence a classe negativa e que foi classificada incorretamente como positiva.
- Falso negativo (FN): instância que pertence a classe positiva e que foi classificada incorretamente como negativa.

# Accuracy (Metz, 1978)

- *Accuracy*: taxa correta de classificação em relação a todos os exemplos.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

# Precision, Recall (Davis e Goadrich, 2006)

- *Precision*: Dentre todos os exemplos classificados como positivos, quais realmente são positivos ?

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

- *Recall* ou *True Positive Rate* - *TPR*: Dentre os exemplos classificados como positivos (que realmente são positivos) e os classificados como negativos (mas que são positivos), qual a taxa de exemplos positivos classificados corretamente?

$$Recall = \frac{TP}{TP + FN} \quad (5)$$



# Precision, Recall (Davis e Goadrich, 2006)

- *Precision*: Dentre todos os exemplos classificados como positivos, quais realmente são positivos ?

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

- *Recall* ou *True Positive Rate* - *TPR*: Dentre os exemplos classificados como positivos (que realmente são positivos) e os classificados como negativos (mas que são positivos), qual a taxa de exemplos positivos classificados corretamente?

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

# stratified-k-fold-cross-validation

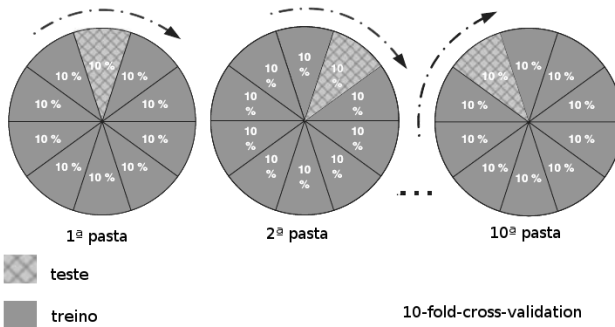


Imagem: k-fold-cross-validation. Adaptado de (Olson e Delen, 2008).

# Experimentos

- Para cada par de atributos selecionados utilizou-se *stratified-10-fold-cross-validation* (repetido dez vezes) juntamente com o *ensemble DECORATE* a fim de verificar se é possível generalizar pelo menos um modelo.
- Em cada verificação construiu-se um modelo com diferentes frações do dataset, variando de 16 instâncias a 59 instâncias e, construiu-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de *recall*.
- Verificou a partir de que ponto esta curva tinha os melhores valores de *recall*, *precision* e *accuracy*.
- Realizou um procedimento de geração de modelo final a partir desse melhor ponto.

# Experimentos

- Para cada par de atributos selecionados utilizou-se *stratified-10-fold-cross-validation* (repetido dez vezes) juntamente com o *ensemble DECORATE* a fim de verificar se é possível generalizar pelo menos um modelo.
- Em cada verificação construiu-se um modelo com diferentes frações do dataset, variando de 16 instâncias a 59 instâncias e, construiu-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de *recall*.
- Verificou a partir de que ponto esta curva tinha os melhores valores de *recall*, *precision* e *accuracy*.
- Realizou um procedimento de geração de modelo final a partir desse melhor ponto.

# Experimentos

- Para cada par de atributos selecionados utilizou-se *stratified-10-fold-cross-validation* (repetido dez vezes) juntamente com o *ensemble DECORATE* a fim de verificar se é possível generalizar pelo menos um modelo.
- Em cada verificação construiu-se um modelo com diferentes frações do dataset, variando de 16 instâncias a 59 instâncias e, construiu-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de *recall*.
- Verificou a partir de que ponto esta curva tinha os melhores valores de *recall*, *precision* e *accuracy*.
- Realizou um procedimento de geração de modelo final a partir desse melhor ponto.

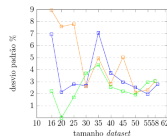
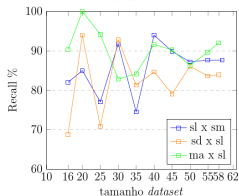
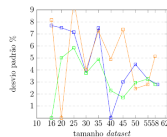
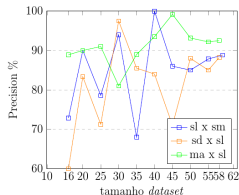
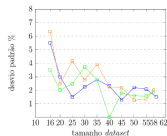
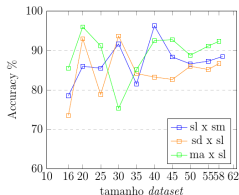
# Experimentos

- Para cada par de atributos selecionados utilizou-se *stratified-10-fold-cross-validation* (repetido dez vezes) juntamente com o *ensemble DECORATE* a fim de verificar se é possível generalizar pelo menos um modelo.
- Em cada verificação construiu-se um modelo com diferentes frações do dataset, variando de 16 instâncias a 59 instâncias e, construiu-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de *recall*.
- Verificou a partir de que ponto esta curva tinha os melhores valores de *recall*, *precision* e *accuracy*.
- Realizou um procedimento de geração de modelo final a partir desse melhor ponto.

# Experimentos

- Para cada par de atributos selecionados utilizou-se *stratified-10-fold-cross-validation* (repetido dez vezes) juntamente com o *ensemble DECORATE* a fim de verificar se é possível generalizar pelo menos um modelo.
- Em cada verificação construiu-se um modelo com diferentes frações do dataset, variando de 16 instâncias a 59 instâncias e, construiu-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de *recall*.
- Verificou a partir de que ponto esta curva tinha os melhores valores de *recall*, *precision* e *accuracy*.
- Realizou um procedimento de geração de modelo final a partir desse melhor ponto.

# Resultados: 10×stratified-10-fold-cross-validation





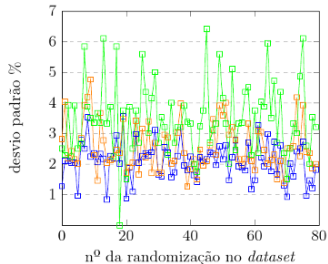
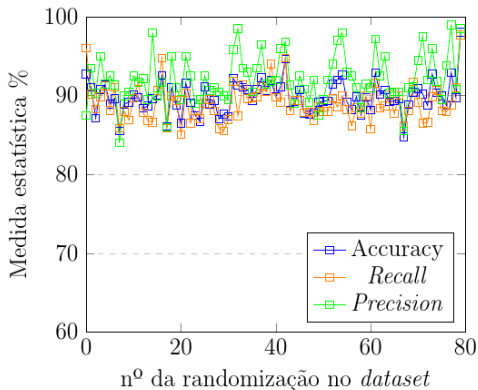
# Resultados: curva de aprendizado selecionada

| $n^{\circ}i$ | 40          | 45          | 50          | 55          | 58          |
|--------------|-------------|-------------|-------------|-------------|-------------|
| <b>A, DP</b> | 92.5, 0     | 92.75, 1.78 | 88.8, 1.6   | 91.13, 1.56 | 92.33, 1.89 |
| <b>P, DP</b> | 93.5, 2.29  | 99.17, 1.71 | 93.17, 2.93 | 92.17, 3.25 | 92.5, 2.81  |
| <b>R, DP</b> | 91.69, 2.56 | 90.25, 2.21 | 86.33, 1.91 | 89.71, 2.97 | 92.05, 3.06 |

**Tabela:** Valores das curvas “ma x sl” da Figura 15 a partir do *dataset* com 40 instâncias. A (*Accuracy*), P (*Precision*), R (*Recall*), DP (Desvio Padrão), i (instâncias). Os valores A, P, R e DV estão em %.

$$(accuracy \geq (92.75 - 1.78)) \wedge (precision \geq (99.17 - 1.71)) \wedge (recall \geq 90.25) \quad (6)$$

# Resultados: geração do modelo final



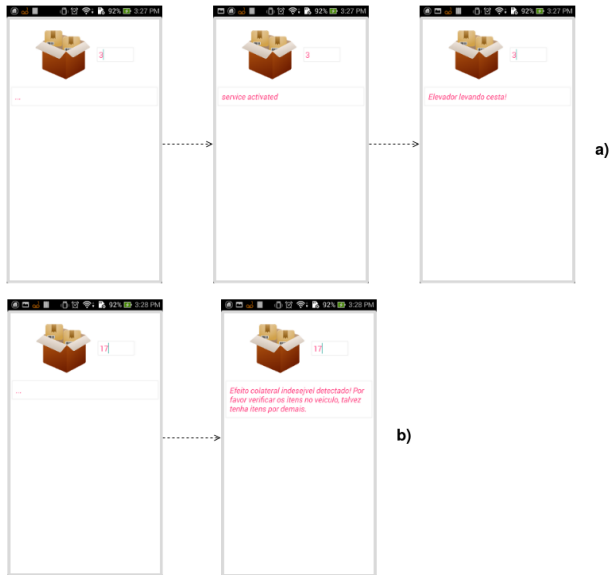
**Imagem:** *stratified-10-fold-cross-validation* (repetido dez vezes). 45 instâncias foram escolhidas randomicamente em cima do *dataset* de 59 instâncias. Procedimento repetido até que satisfizesse a expressão 6.

# Resultados: geração do modelo final

|                    |       | <b>Desvio Padrão %</b> |
|--------------------|-------|------------------------|
| <b>Accuracy %</b>  | 98.05 | 1.81                   |
| <b>Precision %</b> | 98.5  | 3.2                    |
| <b>Recall %</b>    | 97.67 | 2                      |

**Tabela:** Valores da validação do modelo final.

# Resultados: Implantação do modelo final



# Conteúdo

## 1 Introdução

## 2 Proposta

- Cenário Levar compras
- Método de detecção
  - Geração da massa de dados
  - Seleção dos atributos
  - Método ensemble DECORATE
- Implantação do método proposto

## 3 Validação

- Método avaliativo stratified-k-fold-cross-validation
- Experimentos
- Resultados e discussão
  - Modelos
  - Implantação

## 4 Conclusão e trabalhos futuros

# Conclusão

- Foi construído o cenário “Levar compras” no HNS.
- A composição de dispositivos provocou efeitos colaterais indesejáveis no cenário “Levar compras”.
- Obteve-se um modelo baseado no *ensemble* DECORATE com alto grau de generalização em um *dataset* de treino com 45 instâncias com os valores de *Accuracy* = 98.05%, *Precision* = 98.5% e *Recall* = 97.67%.
- Modelo foi implantado no FIM do HNS, desta forma ficou apto a realizar detecção inteligente de efeitos colaterais indesejáveis.

# Conclusão

- Foi construído o cenário “Levar compras” no HNS.
- A composição de dispositivos provocou efeitos colaterais indesejáveis no cenário “Levar compras”.
- Obteve-se um modelo baseado no *ensemble* DECORATE com alto grau de generalização em um *dataset* de treino com 45 instâncias com os valores de *Accuracy* = 98.05%, *Precision* = 98.5% e *Recall* = 97.67%.
- Modelo foi implantado no FIM do HNS, desta forma ficou apto a realizar detecção inteligente de efeitos colaterais indesejáveis.

# Conclusão

- Foi construído o cenário “Levar compras” no HNS.
- A composição de dispositivos provocou efeitos colaterais indesejáveis no cenário “Levar compras”.
- Obteve-se um modelo baseado no *ensemble* DECORATE com alto grau de generalização em um *dataset* de treino com 45 instâncias com os valores de *Accuracy* = 98.05%, *Precision* = 98.5% e *Recall* = 97.67%.
- Modelo foi implantado no FIM do HNS, desta forma ficou apto a realizar detecção inteligente de efeitos colaterais indesejáveis.



# Conclusão

- Foi construído o cenário “Levar compras” no HNS.
- A composição de dispositivos provocou efeitos colaterais indesejáveis no cenário “Levar compras”.
- Obteve-se um modelo baseado no *ensemble* DECORATE com alto grau de generalização em um *dataset* de treino com 45 instâncias com os valores de *Accuracy* = 98.05%, *Precision* = 98.5% e *Recall* = 97.67%.
- Modelo foi implantado no FIM do HNS, desta forma ficou apto a realizar detecção inteligente de efeitos colaterais indesejáveis.

# Trabalhos futuros

- Realizar estudo sobre algoritmos de seleção de atributos e aplicar os algoritmos no cenário “Levar compras”, realizando comparações dos resultados obtidos com o os resultados deste trabalho.
- Criar cenário mais complexo, com pouca ou nenhuma intervenção humana e, realizar os mesmos testes anteriores.
- Realizar estudos sobre aprendizado não supervisionado e prover metodologia para realizar classificação de efeitos colaterais indesejáveis e, realizar os mesmos testes e comparações nos dois cenários anteriores.


# Trabalhos futuros


- Realizar estudo sobre algoritmos de seleção de atributos e aplicar os algoritmos no cenário “Levar compras”, realizando comparações dos resultados obtidos com o os resultados deste trabalho.
- Criar cenário mais complexo, com pouca ou nenhuma intervenção humana e, realizar os mesmos testes anteriores.
- Realizar estudos sobre aprendizado não supervisionado e prover metodologia para realizar classificação de efeitos colaterais indesejáveis e, realizar os mesmos testes e comparações nos dois cenários anteriores.


# Trabalhos futuros

- Realizar estudo sobre algoritmos de seleção de atributos e aplicar os algoritmos no cenário “Levar compras”, realizando comparações dos resultados obtidos com o os resultados deste trabalho.
- Criar cenário mais complexo, com pouca ou nenhuma intervenção humana e, realizar os mesmos testes anteriores.
- Realizar estudos sobre aprendizado não supervisionado e prover metodologia para realizar classificação de efeitos colaterais indesejáveis e, realizar os mesmos testes e comparações nos dois cenários anteriores.


# Referências I


 ARORA, R.; SUMAN, S. Comparative analysis of classification algorithms on different datasets using weka. *International Journal of Computer Applications*, v. 54, n. 13, 2012.

 BARRETT, J. *The Internet of Things - where the Web and the physical world will meet. TED x CIT*, x = independently organized TED event. 2012. Disponível em: <<https://www.youtube.com/watch?v=QaTlt1C5R-M>>.

 CALDER, M. et al. Feature interaction: a critical review and considered forecast. *Computer Networks*, v. 41, p. 115–141, 2003. Disponível em: <<http://eprints.gla.ac.uk/2874/1/feature1calder.pdf>>.


# Referências II


 CHANDRAKANTH, S. et al. Internet of things. *International Journal of Innovations and Advancement in Computer Science IJIACS*, v. 3, October 2014.


 DAVIS, J.; GOADRIC, M. The relationship between precision-recall and roc curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York, NY, USA: ACM, 2006. (ICML '06), p. 233–240. ISBN 1-59593-383-2. Disponível em: <<http://doi.acm.org/10.1145/1143844.1143874>>.


 ENTERPRISE, I. D. N.; MICRO, R. I. G.; NANOSYSTEMS. *Internet of Things in 2020: Roadmap for the Future*. 2008.

# Referências III


 FIRDAUSI, I. et al. Analysis of machine learning techniques used in behavior-based malware detection. In: *Advances in Computing, Control and Telecommunication Technologies (ACT), 2010 Second International Conference on*. [S.l.: s.n.], 2010. p. 201–203.


 HALL, M. et al. The weka data mining software: An update. *SIGKDD Explorations*, v. 11, 2009.


 KARTHIKEYAN, T.; THANGARAJU, P. Analysis of classification algorithms applied to hepatitis patients. *International Journal of Computer Applications*, v. 62, n. 15, 2013.


 MELVILLE, P.; MOONEY, R. J. Constructing diverse classifier ensembles using artificial training examples. In: *Eighteenth International Joint Conference on Artificial Intelligence*. [S.l.: s.n.], 2003. p. 505–510.

# Referências IV

 MELVILLE, P.; MOONEY, R. J. Creating diversity in ensembles using artificial data. *Information Fusion: Special Issue on Diversity in Multiclassifier Systems*, 2004. Submitted.


 METZ, C. Basic principles of roc analysis. *Seminars in Nuclear Medicine*, v. 8, n. 4, p. 283–298, 1978. ISSN 0001-2998. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0001299878800142>>.


 NAKAMURA, M.; IKEGAMI, K.; MATSUMOTO, S. Considering impacts and requirements for better understanding of environment interactions in home network services. *Computer Networks*, Elsevier, v. 57, p. 2442–2453, 2013.


 NHLABATSI, A.; LANEY, R.; NUSEIBEH, B. Feature interaction: the security threat from within software systems. *Progress in Informatics*, n. 5, p. 75–89, 2008. Disponível em: <[http://www.nii.jp/pi/n5/5\\_75.pdf](http://www.nii.jp/pi/n5/5_75.pdf)>.



# Referências V

 OLSON, D.; DELEN, D. *Advanced Data Mining Techniques*. [S.l.]: Springer-Verlag Berlin Heidelberg, 2008. ISBN 978-3-540-76916-3.

 PAUTASSO, C. Restful web services: Principles, patterns, emerging technologies. In: *Springer Science+Business Media*. New York: [s.n.], 2014. Disponível em: <<http://vis.uky.edu/~cheung/courses/ee586/papers/Pautasso2014.pdf>>.

 SUNDMAEKER, H. et al. (Ed.). *Vision and Challenges for Realising the Internet of Things*. [S.l.]: European Union, 2010. ISBN 978-92-79-15088-3.

# Obrigado Pela Atenção!

# Perguntas?