

# Uso de Redes Neurais e Informação Mútua na Análise de Dados Eleitorais

Francisco Morgani Fatore

30 de julho de 2012

## Resumo

Este trabalho buscou analisar os resultados das últimas eleições estaduais e nacionais realizadas no Brasil. Buscou-se descobrir e entender os padrões que levam um candidato a ser eleito. Este objetivo foi alcançado por meio de técnicas de aprendizado de máquina e de teoria da informação. Uma rede neural foi capaz de prever com uma alta taxa de acerto quando um candidato é eleito com base em suas informações pessoais. Como não se sabia ao certo quais informações seriam relevantes para este tipo de análise, utilizou-se a informação mútua entre os atributos para encontrar os fatores que realmente importavam para o resultado das eleições.

## 1 Introdução

A capacidade de armazenamento de dados alcançada pelo atual desenvolvimento tecnológico viabilizou a muitas empresas e instituições registrar dados relativos aos seus negócios em busca de obter melhores resultados e respostas de seus serviços. Porém, existe um longo caminho entre o armazenamento dos dados e a extração e interpretação de informações de interesse.

Empresas como Google e Facebook dedicam grande parte de seus esforços no sentido de explorar esses dados e dependem desta análise para fornecer uma publicidade direcionada aos seus clientes. A exploração de grandes bases de dados não precisa estar vinculada a algum interesse comercial, mas como essa atividade é muito dispendiosa acaba ocorrendo principalmente com esse pretexto ou dentro de algum ambiente de pesquisa.

Existe uma imensa quantidade de dados sem um aparente interesse financeiro que pode oferecer à população recursos para compreender melhor a sociedade em diferentes aspectos. O governo brasileiro, por exemplo, tem trabalhado junto à OGP<sup>1</sup> (Open Government Partnership) no sentido de oferecer uma maior transparência de suas atividades à população por meio da disponibilização de dados públicos. Exemplos desse esforço são os portais

eletrônicos Portal da Transparência<sup>2</sup> e Portal Brasileiro de Dados Abertos<sup>3</sup>.

Contudo, a iniciativa do governo é muito recente, formalizada no final de 2011 pelo Decreto de 15 de setembro de 2011 [1] e pela Lei de Acesso à Informação Pública [2], ainda requer um amadurecimento em questões de organização e aproveitamento dos dados. Hoje, o governo se limita a disponibilizar os dados em sua forma bruta, ou seja, sem suporte a análise exploratória com intuito de extrair informações de interesse e ajudar na sua interpretação.

O Portal da Transparência, mencionado anteriormente, possui uma vasta e variada base de dados, porém, em sua maioria, eles se encontram no formato de planilhas que são de difícil interpretação. Já o segundo portal mencionado, o Portal de Dados Abertos, disponibiliza os dados em formatos que podem ser tratados em larga escala, entretanto, não possui um registro tão vasto e variado como o do Portal da Transparência.

Outro recurso fornecido pelo governo, por meio do TSE (Tribunal Superior Eleitoral), trata-se do Repositório de Dados Eleitorais<sup>4</sup>. Este recurso conta com uma base de dados referente a candidatos, eleitorado, resultados e prestação de contas. Os registros contêm informações de eleições municipais, estaduais e nacionais de 1994 até 2010 e encontram-se em formatos adequados para o processamento computacional, configurando uma fonte valiosa para entendimento dos mecanismos de funcionamento da política nacional.

Este trabalho buscou utilizar técnicas de aprendizado de máquina e teoria da informação para fornecer recursos para um aprofundamento no estudo dos dados eleitorais contidos no repositório de dados do TSE. Mais especificamente construiu-se um modelo computacional baseado em redes neurais que foi capaz de avaliar as chances de um candidato ser eleito com base nas informações de sua candidatura.

A rede neural apresentou sucesso satisfatório em suas previsões. Sucesso que foi viabilizado pela compreensão

<sup>1</sup><http://www.opengovpartnership.org/>

<sup>2</sup><http://www.portaltransparencia.gov.br/>

<sup>3</sup><http://beta.dados.gov.br/>

<sup>4</sup><http://www.tse.jus.br/eleicoes/repositorio-de-dados-eleitorais>

das interdependências entre os atributos da base de dados. Este entendimento sobre os dados foi extraído a partir do uso da informação mútua entre os atributos.

Na próxima seção apresenta-se alguns trabalhos que se relacionam com esse trabalho, seja pelo uso de técnicas semelhantes ou por motivações concorrentes.

## 2 Trabalhos Relacionados

O trabalho realizado por Battiti [3] é uma das bases teóricas deste trabalho pois trata exatamente de uma das questões mais importantes na implementação de uma rede neural, a seleção dos atributos de entrada da rede. Battiti defende um método para seleção de atributos baseado em informação mútua.

Nessa questão de seleção de atributos, o trabalho realizado por Peng, Long e Ding [4] é muito interessante, pois aborda a visualização das interdependências entre atributos utilizando informação mútua. Os autores atentam para a importância da etapa de pré-processamento e ressaltam que o entendimento da interdependência entre os atributos é crucial para a seleção de atributos.

Existe na literatura outros trabalhos na mesma vertente motivacional deste trabalho. Vieira e Lobo [5], por exemplo, utilizaram dados das eleições municipais de Portugal no ano de 1999 em uma comparação entre as técnicas de agrupamento hierárquico e as redes neurais de Kohonen, conhecidas como *Self-Organizing Maps* (SOM) [6]. O resultado com ambos métodos mostrou a formação de três grupos: um apontando o Partido Socialista (PS) como partido mais votado, sendo este grupo bem maior que os demais; um grupo dominado por partidos de direita; e outro grupo que se concentra na região de Alentejo, sendo dominado por partidos de esquerda.

Outro trabalho que pode ser mencionado é o de Carneiro e Almeida [7], o qual busca observar como as disputas eleitorais nos municípios são influenciadas pela disputa nas esferas estaduais e nacional. Entre as conclusões apresentadas neste trabalho, destacam-se: a votação para governador é a variável que causa maior efeito nas eleições de deputado federal e estadual; e a votação para prefeito é a que tem maior efeito para a eleição de vereadores. Os autores afirmam que é extremamente difícil a observação da organização partidária, devido à grande quantidade de municípios em que se encontram partidos políticos.

Um último exemplo que vale ser citado é o de Gormley e Murphy [8], o qual busca analisar a influência de fatores sociais nos resultados eleitorais. Os autores utilizam modelos de mistura para evidenciar os grupos de votação no eleitorado irlandês das eleições presidenciais de 1997 e determinar as características nas preferências de votação desses blocos. Como resultado, os autores concluem que a idade e a satisfação com o governo determinam a forma-

ção dos grupos de votação no caso estudado.

## 3 Metodologia

Este trabalho busca treinar uma rede neural MLP [9] (*Multilayer Perceptron*) com os dados das eleições de um determinado ano e verificar se essa rede é capaz de prever se um candidato foi eleito.

Como já foi mencionado, os dados utilizados neste trabalho são provenientes do repositório eleitoral de dados do tribunal superior eleitoral e isso garante a legitimidade das informações. No entanto, não há garantias que a base de dados é totalmente consistente. De fato, uma análise prévia dos dados revelou uma série de inconsistências na base, como alguns atributos em branco ou certas replicações de dados que sugerem falhas durante a etapa de inserção dos dados. Logo, foi necessário realizar um pré-processamento nos dados para diagnosticar essas falhas e removê-las.

O repositório de dados contém informações relativas a *Eleitorado, Candidatos, Resultados e Prestação de Contas*. Devido ao foco deste trabalho escolheu-se por trabalhar com os dados relativos a candidatos. Dentro dessa categoria existem as subcategorias *Informações de Candidatura, Bens, Legendas e Vagas*. Decidiu-se trabalhar somente com os dados de informações de candidatura pois continham atributos suficientes para a análise proposta.

A fase de pré-processamento consistiu em criar um banco de dados a partir dos arquivos obtidos no repositório de dados. Ao término desta etapa o banco de dados se encontra na situação descrita pela tabela 1.

Alguns desses atributos não são relevantes para a análise ou não se encontram em um formato adequado para o processamento. Deste modo, após uma etapa de seleção manual dos atributos obtém-se a tabela 2.

Os atributos de 1 a 7 são valores *discretos* equivalentes aos identificadores dos campos na base de dados. Os atributos de 8 a 11 são valores *contínuos* que devem ser divididos em intervalos para se adequarem ao processamento. Essa divisão em intervalos foi realizada por meio da criação de um histograma, onde cada valor contínuo foi representado pelo intervalo do histograma o qual pertence. Métodos mais adequados poderiam ser utilizados, como por exemplo o proposto por Darbellay e Vajda [10]. O valor 12 é o *label* que indica se o candidato foi eleito ou não. Todos os atributos foram normalizados entre 0 e 1 para garantir uma maior uniformidade nos cálculos.

Antes de processar os dados na rede neural deve-se analisar as possíveis relações entre os atributos da base de dados, de modo a escolher os que melhor representam os dados. Isso será feito utilizando a informação mútua entre os atributos. A análise da interdependência entre os atributos permite realizar uma seleção de atributos.

Tabelas	
Eleição	Ano Número do Turno Descrição Unidade Federativa (UF) Unidade Eleitoral (UE) Cargo
Candidato	Nome Data Nascimento UF Nascimento Cidade Nascimento Título Eleitoral Nacionalidade
Situação Candidato	Ocupação Idade Grau Instrução Estado Civil
Candidatura	Número de Urna Nome de Urna Situação Despesa Máxima Resultado
Partido	Número Sigla Nome
Legenda	Código Sigla Composição Nome

Tabela 1: Tabelas do banco de dados criado na etapa de pré-processamento dos dados.

Atributos	
1	Cidade Natal
2	Estado Natal
3	Nacionalidade
4	Sexo
5	Partido Político
6	Ocupação Atual
7	Estado Civil
8	Nível do Cargo Almejado
9	Grau de Escolaridade
10	Idade na Eleição
11	Despesas de Campanha Estipulada
12	Resultado da Eleição

Tabela 2: Atributos selecionados manualmente.

Realizada a seleção de atributos, resta apenas preparar os dados para serem processados pela rede neural. Isso é necessário pois os valores discretos (atributos 1 a 7) não apresentam um fluxo de dados, ou seja, não se pode, por

exemplo, dizer que uma cidade representada pelo código  $x$  esta geograficamente próxima de uma representada pelo código  $x + 1$ . Para contornar essa situação basta ortogonalizar os valores, por exemplo se o valor sexo era representado por 0 e 1, passa a ser representando por  $[1, 0]$  e  $[0, 1]$ . Os valores contínuos já foram processados na etapa anterior e não requerem mais nenhum tipo de tratamento para serem processados pela rede. O critério de parada e a taxa de aprendizado da MLP serão definidos empiricamente de acordo com os recursos computacionais disponíveis.

A seção 4 apresenta os resultados obtidos por esta abordagem proposta.

### 3.1 Informação Mútua

A informação mútua é uma medida estatística que indica a dependência ente duas variáveis. Formalmente, a informação mútua entre duas variáveis  $X$  e  $Y$  pode ser definida para o caso discreto como [4]:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right), \quad (1)$$

onde  $p(x, y)$  é a função de distribuição de probabilidade conjunta de  $X$  e  $Y$ , e  $p(x)$   $p(y)$  são as funções de probabilidade marginais de  $X$  e  $Y$  respectivamente.

Trata-se de uma medida que é aplicável a variáveis contínuas, discretas e categóricas, o que a torna uma ferramenta preciosa para a análise proposta por este trabalho. Além disso, essa medida apresenta algumas vantagens sobre outros métodos de avaliação de dependência entre atributos, como por exemplo a correlação. A correlação é utilizada para capturar apenas dependências lineares, e sua aplicação é restrita a atributos quantitativos.

## 4 Resultados

Trabalhou-se com os dados referentes às eleições estaduais e nacionais de 2010. Havia ao todo, 19210 elementos na base de dados, porém 286 desses continham atributos em branco ou inválidos e por isso foram descartados.

Com o objetivo de analisar a relação entre os atributos da base de dados, construiu-se uma matriz (1) com as informações mútuas de todos os atributos entre si. Trata-se de uma matriz  $m \times m$ , em que  $m$  indica o número de atributos e que  $m(i, j)$  indica a informação mútua do atributo  $i$  com o atributo  $j$ .

Uma análise de um ponto de vista geral, realizada a partir da matriz de informação mútua é capaz de fornecer interpretações interessantes dos dados. Por exemplo, nota-se que o atributo sexo (4) influencia praticamente todos os outros atributos, inclusive o resultado da eleição (12). Essa suposição pode ser confirmada estatisticamente, pois

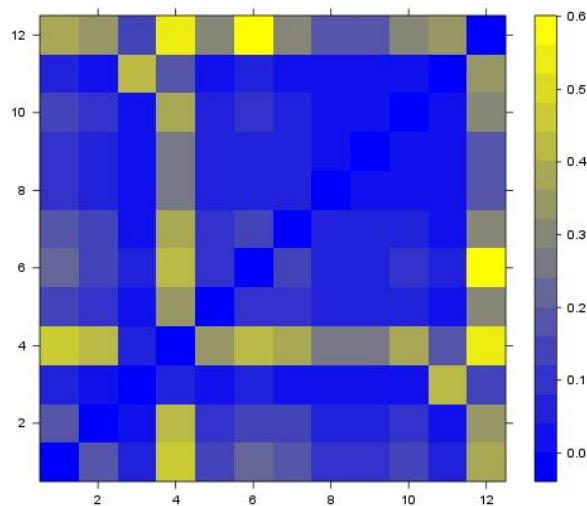


Figura 1: Matriz dos valores de informação mútua para todos os atributos. Valores maiores representam maior dependência entre os atributos. A diagonal deve ser ignorada.

dos 1424 candidatos eleitos em 2010, apenas 168 eram do sexo feminino, ou seja, mais de 88% dos cargos foram ocupados por pessoas do sexo masculino.

Nota-se também que o resultado da eleição apresenta relação com a maioria dos atributos. Para uma melhor análise criou-se um gráfico (2) com as informações mútuas dos atributos em relação ao resultado da eleição.

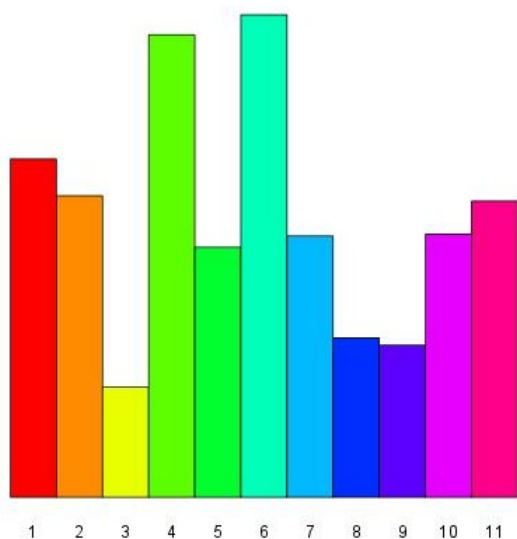
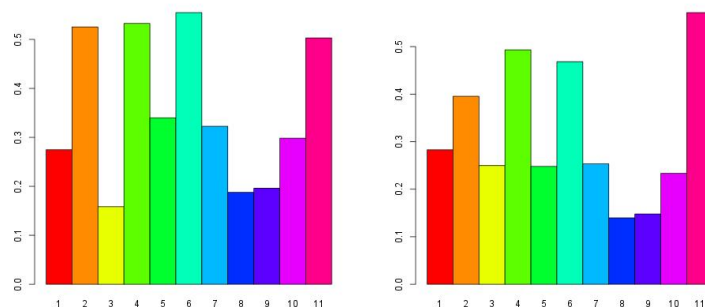


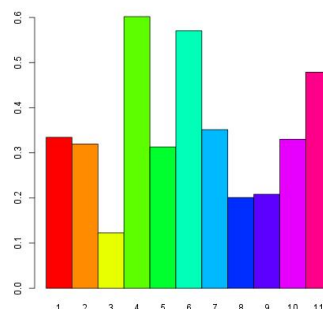
Figura 2: Gráfico da informação mútua de todos atributos em relação ao resultado da eleição.

Observa-se que os atributos correspondentes ao sexo (4) e a ocupação (6) do candidato são os que mais influenciam no resultado da eleição. Existe também uma certa influência causada pelo local de nascimento do candidato (1 e 2). Um fato interessante é que o valor estipulado para despesas na campanha (11) não apresenta a maior relação com o resultado da eleição, contrariando o senso comum.

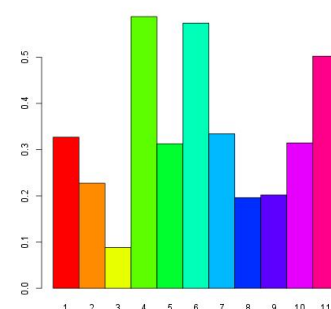


(a) RS

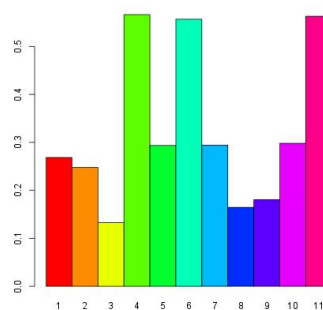
(b) SP



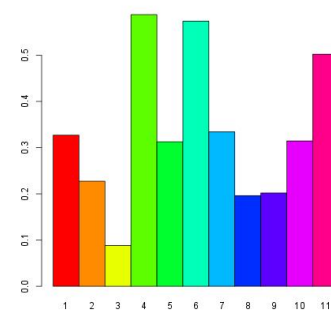
(c) BA



(d) MA



(e) GO



(f) AM

Figura 3: Informação mútua calculada individualmente para os estados.

Uma hipótese que pode ser avaliada é se existe uma variação desses valores se for realizada uma análise mais detalhada dos dados. Uma possibilidade é realizar uma análise

lise individual para cada estado. Como o Brasil apresenta diversos estados optou-se por escolher somente os mais representativos de cada região.

Observa-se pela figura 3 que alguns estados apresentam uma distribuição distinta de valores de informação mútua e diferentes atributos influenciam no resultado da eleição. Nota-se por exemplo que as eleições no Rio Grande do Sul são mais influenciadas pelo estado natal do candidato do que em outras partes do Brasil. Nota-se quem em São Paulo o valor definido para despesas de campanha é o atributo mais significativo. O restante dos estados segue a tendência nacional com uma certa variação em alguns atributos.

Dessa última análise pode-se concluir que o treinamento da rede neural deve ser realizado individualmente para cada estado, devido à variância da informação mútua dos atributos. Assim, uma abordagem seria criar uma rede neural para cada estado, treiná-las com os dados de 2006 e testá-las com os dados de 2010 para verificar sua precisão.

Neste momento é necessário avaliar as questões de desempenho computacional. Supõe-se o processamento da rede no pior caso, ou seja, o caso com o maior número de elementos, que trata-se da eleição no estado de São Paulo. Nesta situação cada elemento é representado por um vetor de 1310 posições. Isso ocorre pois na etapa de ortogonalização, atributos como cidade natal, que possui para este caso 1097 possibilidades, são convertidos em vetores ortogonais, no caso de cidade natal em um vetor de 1097 posições (dimensões). O elevado número de elementos dificulta tanto a convergência da rede neural quanto a qualidade do treinamento obtido. Assim, requer-se uma simplificação dessa representação.

Uma possibilidade para essa simplificação é ignorar o atributo cidade natal (1), pois o atributo estado natal (2) já é capaz de representar essa noção de localidade. Os atributos correspondentes a cidadania (3), nível do cargo almejado (8) e escolaridade (9) não são cruciais pois, de acordo com a figura 2, não apresentam grande relação com o resultado da eleição e por isso podem ser descartados. Assim, no caso do estado de São Paulo por exemplo, há uma redução de 1102 dimensões, resultando em um vetor de 208 dimensões.

No entanto, 208 ainda é um número elevado de dimensões. O atributo ocupação (6) representa 147 dessas dimensões. Uma tentativa para reduzir ainda mais o número de dimensões seria ignorar esse atributo, no entanto como já foi observado, existe uma grande relação entre a ocupação de um candidato e o resultado da eleição. Outra abordagem seria utilizar o senso comum como hipótese para a criação de uma categoria de ocupação, isto é, espera-se que candidatos que se encontram envolvidos com política devem apresentar um desempenho na campanha diferente de candidatos que não estão envolvidos.

Para a criação dessa categoria foram selecionados ma-

nualmente do conjunto de ocupações aquelas que pareciam pertinentes ao grupo de pessoas envolvidas com política. A tabela 3 representa os elementos que compõem esta categoria.

Ocupações
Vereadores
Deputados
Prefeitos
Diplomatas
Governadores
Ministros
Membros do Ministério Público
Ocupantes de cargos em comissões
Senadores

Tabela 3: Categoria de ocupações relacionadas à política.

Uma vez construída essa categoria, basta utilizar um único valor binário para indicar quando o candidato se encontra nesta categoria ou não. Uma análise estatística sobre os dados referentes às eleições no estado de São Paulo revela que 49.66% dos candidatos que obtiveram sucesso em suas eleições pertencem à esta categoria de ocupações. Sabendo que existe na base de dados 209 ocupações cadastradas é possível dizer que esse percentual representa um quanta considerável.

A mesma ideia de criação de uma categoria pode ser utilizada para o atributo estado natal. Usando o senso comum infere-se que um candidato tem maior propensão a ser eleito se candidatando no estado em que nasceu do que em outros estados. Assim o valor estado natal, que é representado por um vetor de 26 posições, passa a ser representado por um único valor booleano que indica se o candidato nasceu no estado em que concorre ou não.

Uma análise estatística pode fundamentar esta última hipótese. Realizando um levantamento das eleições em São Paulo constatou-se que 72.85% das pessoas eleitas nasceram no estado de São Paulo. O gráfico de informação mútua apresentado na figura 3-a mostrou que o Rio Grande do Sul sofria maior influência desta questão de localidade do que outros estados. Um levantamento dos resultados das eleições neste estado revelou que 96.15% dos eleitos nasceram no Rio Grande do Sul.

Com todas as devidas reduções de dimensionalidade atinge-se o número de 38 dimensões para o caso das eleições no estado de São Paulo. A rede neural deve ser capaz de processar essa quantidade de dimensões.

Os parâmetros da rede neural foram definidos empiricamente buscando um equilíbrio entre desempenho computacional e precisão da técnica. O número de neurônios na camada de entrada é o número de atributos de entrada

		Valor Predito		Total
		Eleito	Não Eleito	
Valor Verdadeiro	Eleito/	103	109	P'
	Não Eleito/	0	2568	N'
Total		P	N	

Tabela 4: Matriz de confusão construída a partir dos dados obtidos pela rede neural.

desconsiderando o *label*, ou seja, 37 neurônios. O número de neurônios na camada escondida foi escolhido empiricamente como o valor 17. A camada de saída contém apenas um neurônio, que indica se o candidato foi eleito ou não. O valor da taxa de aprendizado  $\alpha$  foi definido como 0.1. O treinamento da rede termina quando a soma dos erros para todos os valores de entrada é menor do que 1.1. Os dados de entrada são divididos arbitrariamente, de modo que o arquivo de treinamento contenha 10% dos dados e o restante é destinado para o teste da rede.

Ao término do processamento a MLP foi capaz de acertar 96.07% dos resultados das candidaturas. No entanto, somente a taxa de acerto não diz muito sobre a qualidade dos resultados obtidos. Uma matriz de confusão 4 pode auxiliar na análise dos resultados:

Observando os valores da matriz de confusão nota-se que a rede neural foi capaz de distinguir todas as condições no qual um candidato não tem chance de se eleger, pois não existem falsos positivos. No entanto, existe um número considerável de falsos negativos, o que indica uma certa falta de precisão do resultado.

Calculando-se a revocação e a especificidade sobre a matriz de confusão obtém-se os valores 48.58% e 100.0%. A especificidade revela que para o teste realizado a rede foi capaz de garantir que todo candidato predito como eleito realmente foi eleito. No entanto, a rede não apresenta uma alta revocação, ou seja, ela tende a prever um candidato como fracassado mesmo quando este apresenta uma chance de obter sucesso.

## 5 Conclusão

A extração de informações de interesse de grandes bases de dados é uma atividade complexa e trabalhosa. Essa complexidade aumenta ainda mais ao se lidar com conjun-

tos de dados reais que apresentam variedade entre os tipos de atributos. Necessita-se destinar um grande esforço para a atividade de seleção de atributos e neste trabalho esta etapa se mostrou a mais trabalhosa, pois exigiu uma compreensão das características dos atributos da base de dados.

Por meio do uso da informação mútua foi possível inferir uma série de fatores que contribuem para o sucesso de um candidato em sua campanha. Algumas conclusões como a influência do sexo do candidato e seu estado natal no resultado da eleição já eram esperadas. Foi possível também alcançar conclusões não tão óbvias, como por exemplo que o partido de um candidato é tão importante quanto o estado civil do candidato e que as despesas estipuladas para a campanha não são tão importantes como o esperado.

A análise individual dos estados revelou uma série de variações que indicam uma certa dinâmica dos atributos de entrada de acordo com a região do país. São Paulo, por exemplo, recebe uma maior influência do valor de despesas de campanha estipulado do que outros estados. Os candidatos nascidos no Rio Grande do Sul têm uma maior tendência de alcançar sucesso nas eleições neste estado do que candidatos nascidos em qualquer outra parte do país.

Por sua vez, a rede neural apresentou uma altíssima taxa de acerto. Isso pode ser justificado pela correta definição dos atributos de entrada. No entanto, ao observar a matriz de confusão verifica-se que a taxa de acerto não representa com exatidão a qualidade dos resultados obtidos. Devido ao alto desbalanceamento dos dados de entrada a rede acaba rotulando certos candidatos como fracassados mesmo quando eles têm chance de serem eleitos. Esse fato impossibilita garantir que um determinado candidato não seja eleito.

No entanto, para os casos em que a rede determinou que o candidato seria eleito, o fato se mostrou verdadeiro em 100.0% dos casos. Essa conclusão poderia ser utilizada como base de confiança para um candidato, caso ele fosse rotulado como eleito, então ele teria chances reais de obter sucesso em sua campanha.

## Referências

- [1] Brasil. *Decreto de 15 de setembro de 2011*. [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2011-2014/2011/Dsn/Dsn13117.htm](http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Dsn/Dsn13117.htm).
- [2] Brasil. *Lei de Acesso a Informação Pública (Lei 12.527/2011), de 18 de novembro de 2011*. [http://www.presidencia.gov.br/ccivil\\_03/\\_Ato2011-2014/2011/Lei/L12527.htm](http://www.presidencia.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm).

- [3] R Battiti. “Using mutual information for selecting features in supervised neural net learning.” Em: *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* 5.4 (jan. de 1994), pp. 537–50. ISSN: 1045-9227. DOI: 10.1109/72.298224. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18267827>.
- [4] H. Peng, F. Long e C. Ding. “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy.” Em: *IEEE transactions on pattern analysis and machine intelligence* 27.8 (ago. de 2005), pp. 1226–38. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2005.159. URL: <http://www.ncbi.nlm.nih.gov/pubmed/16119262>.
- [5] A. Vieira e V. Lobo. “Agrupamento do padrões por métodos hierárquicos e com Self-Organizing Maps (SOM)”. Em: *XII Jornadas de Classificação e Análise de Dados – JOCLAD*. Lisboa, Portugal, 2006, pp. 6–8.
- [6] T. Kohonen, M. R. Schroeder e T. S. Huang, eds. *Self-Organizing Maps*. 3rd. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2001. ISBN: 3540679219.
- [7] L. P. Carneiro e M. H. T. Almeida. “Definindo a Arena Política Local: Sistemas Partidários Municipais na Federação Brasileira”. Em: *Revista de Ciências Sociais* 51.2 (2008), pp. 403–432.
- [8] I. C. Gormley e T. B. Murphy. “A mixture of experts model for rank data with applications in election studies”. Em: *Annals of Applied Statistics* 2 (2008), pp. 1452–1477. DOI: 10.1214/08-AOAS178.
- [9] D. W. Ruck et al. “The multilayer perceptron as an approximation to a Bayes optimal discriminant function.” Em: *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* 1.4 (jan. de 1990), pp. 296–8. ISSN: 1045-9227. DOI: 10.1109/72.80266. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18282850>.
- [10] G. Darbellay e I. Vajda. “Estimation of the information by an adaptive partitioning of the observation space”. Em: *IEEE Transactions on Information Theory* 45.4 (maio de 1999), pp. 1315–1321. ISSN: 00189448. DOI: 10.1109/18.761290. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=761290>.