

Previsão de Demanda de Passageiros em Sistemas Metroferroviários:

uma abordagem por **Redes Neurais Artificiais**

Vagner Sanches Vasconcelos
vsvasconcelos@gmail.com

Contexto do Trabalho

Demanda de Passageiros

Seleção das variáveis de Entrada

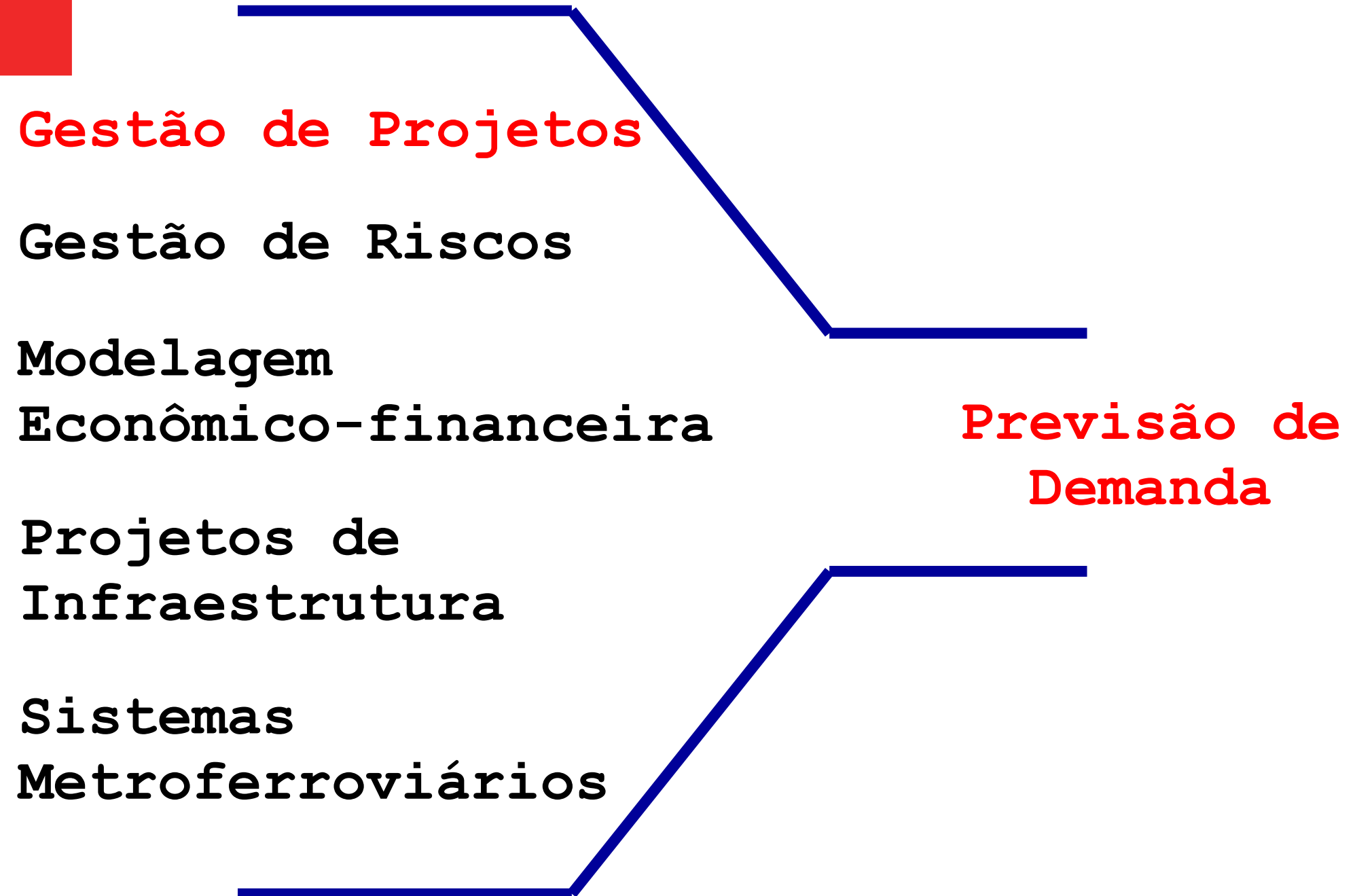
Pré-processamento dos dados

Configuração da Rede Neural

Treinamento, Teste e Validação

Métricas de Desempenho

Agenda



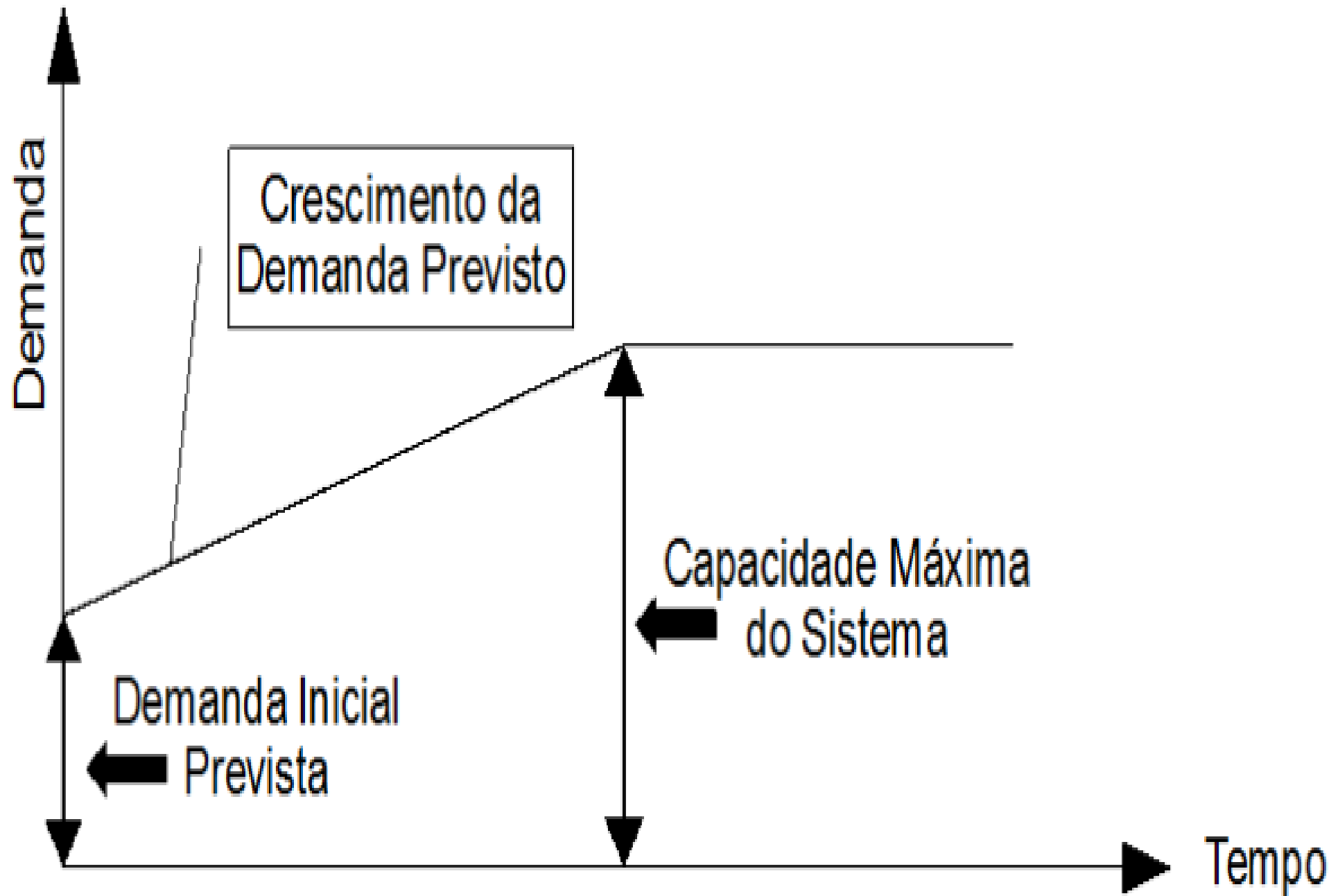
Fatores de risco em projetos de infraestrutura de transporte de passageiros sobre trilhos na modalidade PPP (2013)



Identificação dos principais fatores de risco em projetos de infraestrutura de transporte de passageiros sobre trilhos na modalidade PPP (2014)



Riscos		Linha 3 Metrô RJ	Linha 4 Metrô SP	Expresso Aeroporto CPTM	TAV SP/RJ
Operação	Operação				
	Manutenção				
	Custos mais elevados que os previstos				
	Demanda menor que a estimada				
	Tarifa insuficiente por não haver atualização adequada				
	Verificação de nível de gratuidade superior ao original projetado				
	Desvalorização cambial				



Impacto dos indicadores PIB e Taxa de Desemprego na Demanda de Passageiros da Linha 3 - Vermelha do Metrô-SP (2014)



$\text{Demanda} = f(\text{PIB}, \text{Taxa Desemprego})$

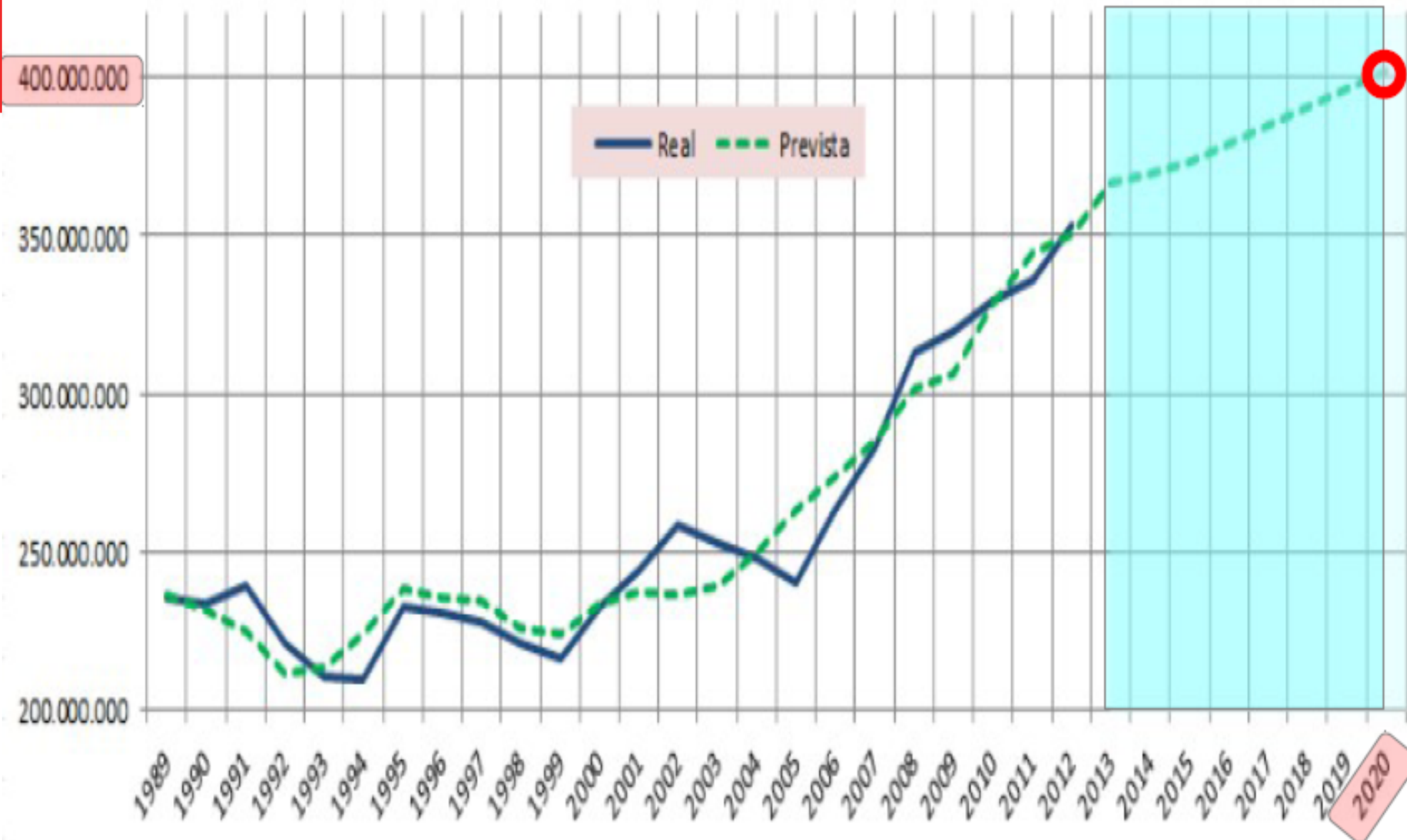
Análise Multivariada

GDP and Unemployment Rate Impact in Demand for Subway Passengers in São Paulo (2015)

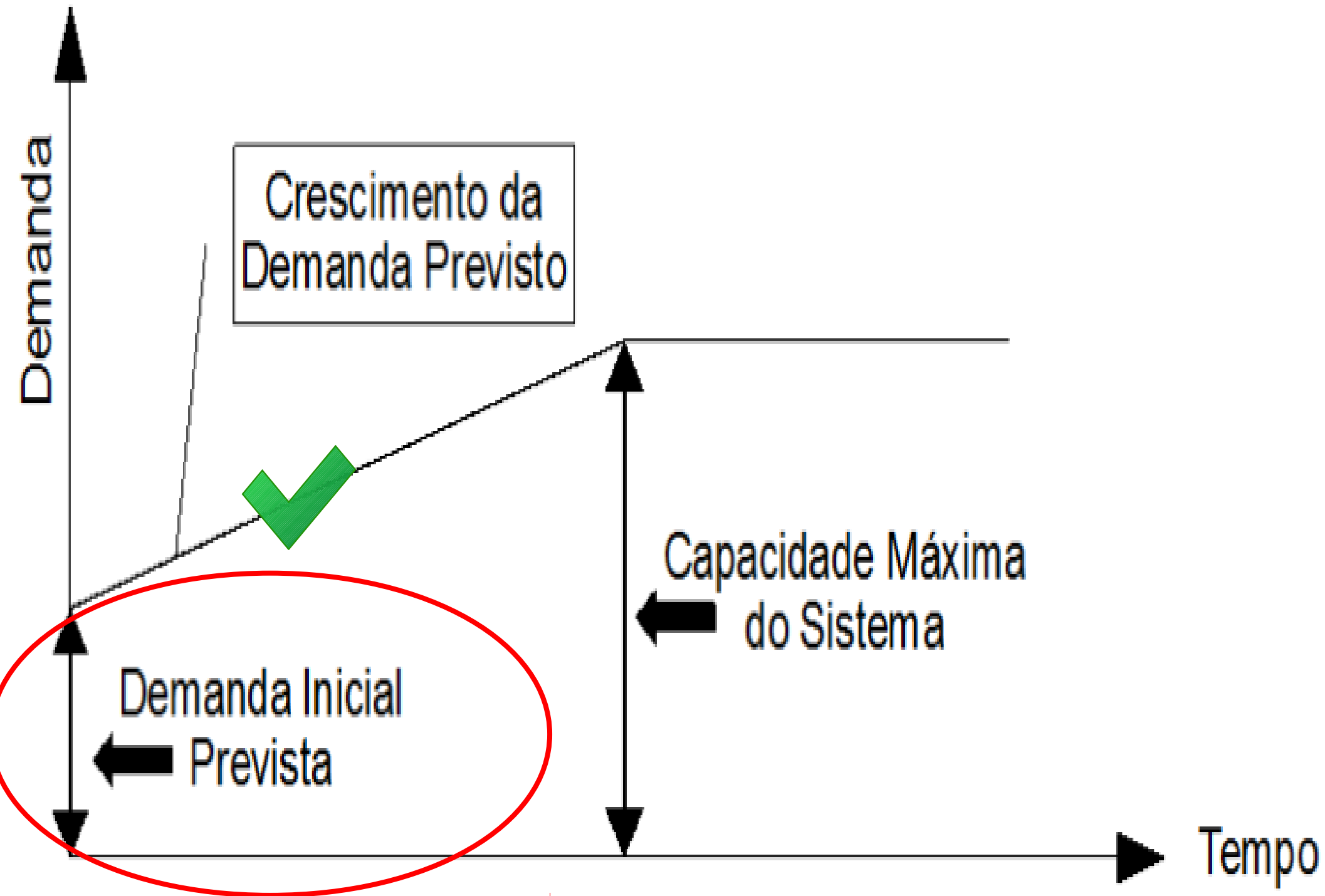


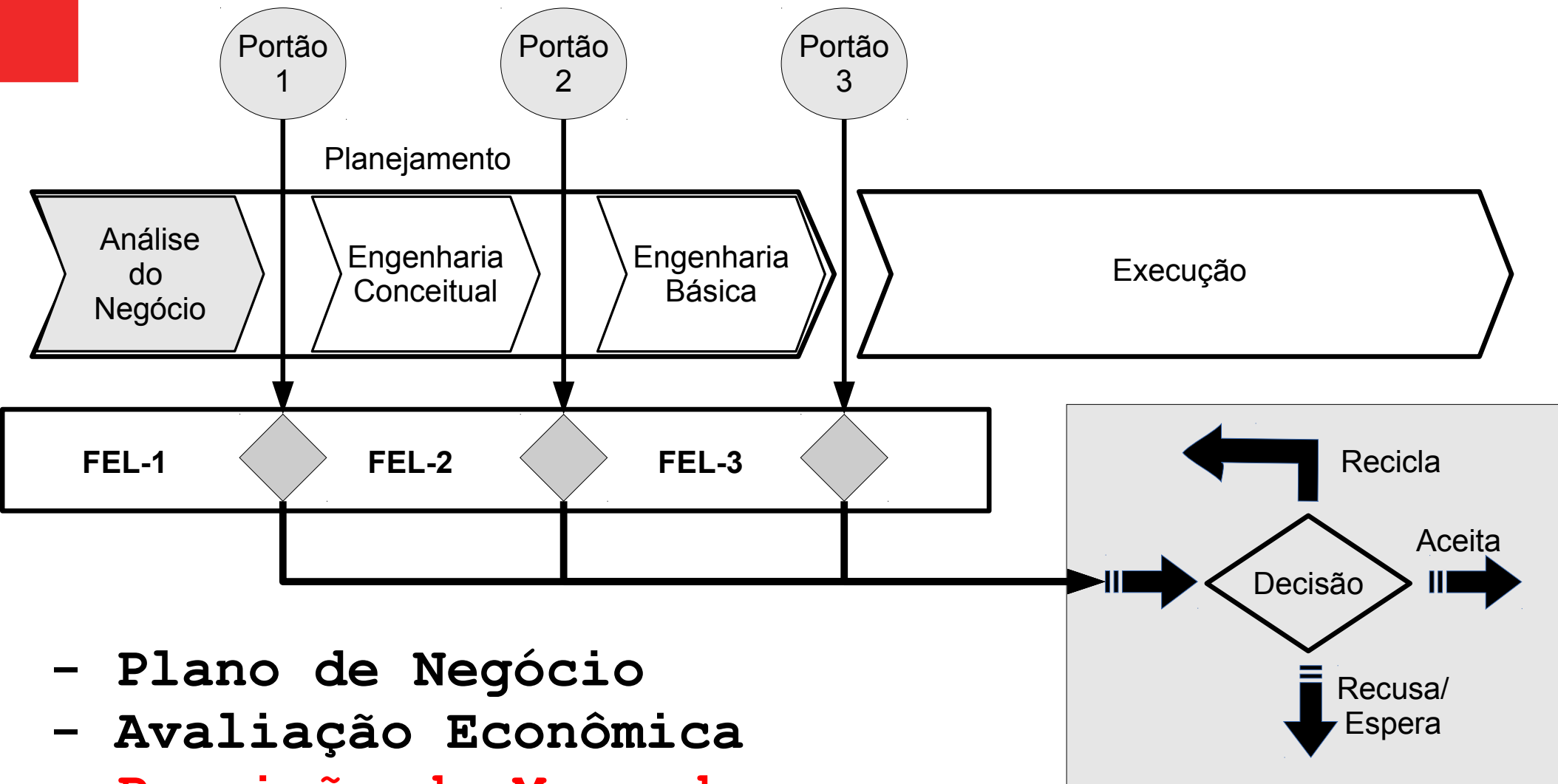
Crescimento da Demanda no Tempo

8



$$Demanda = 271.367.416 + 27,9 * PIB - 4.004.648,6 * Taxa Desemprego + u$$

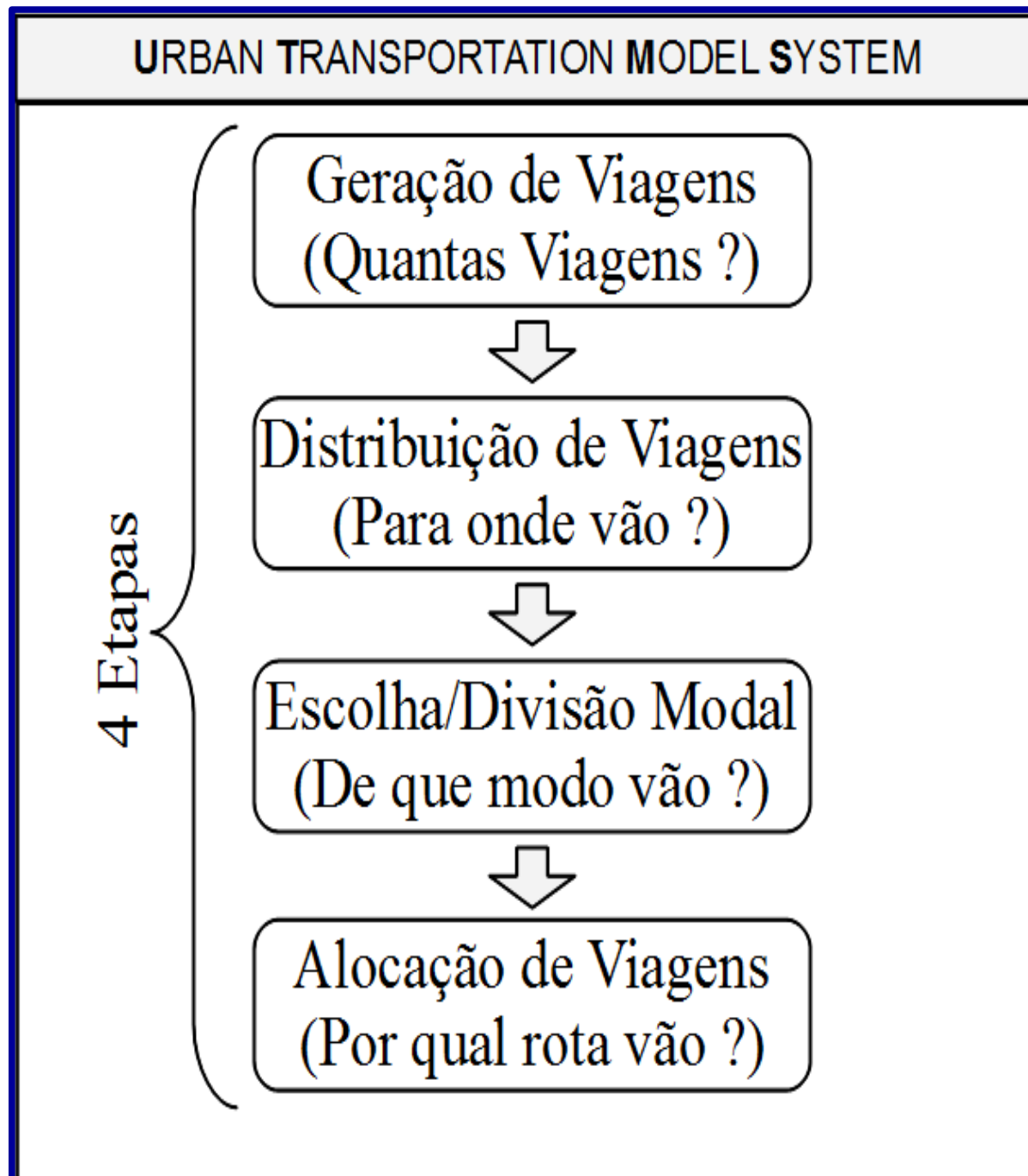




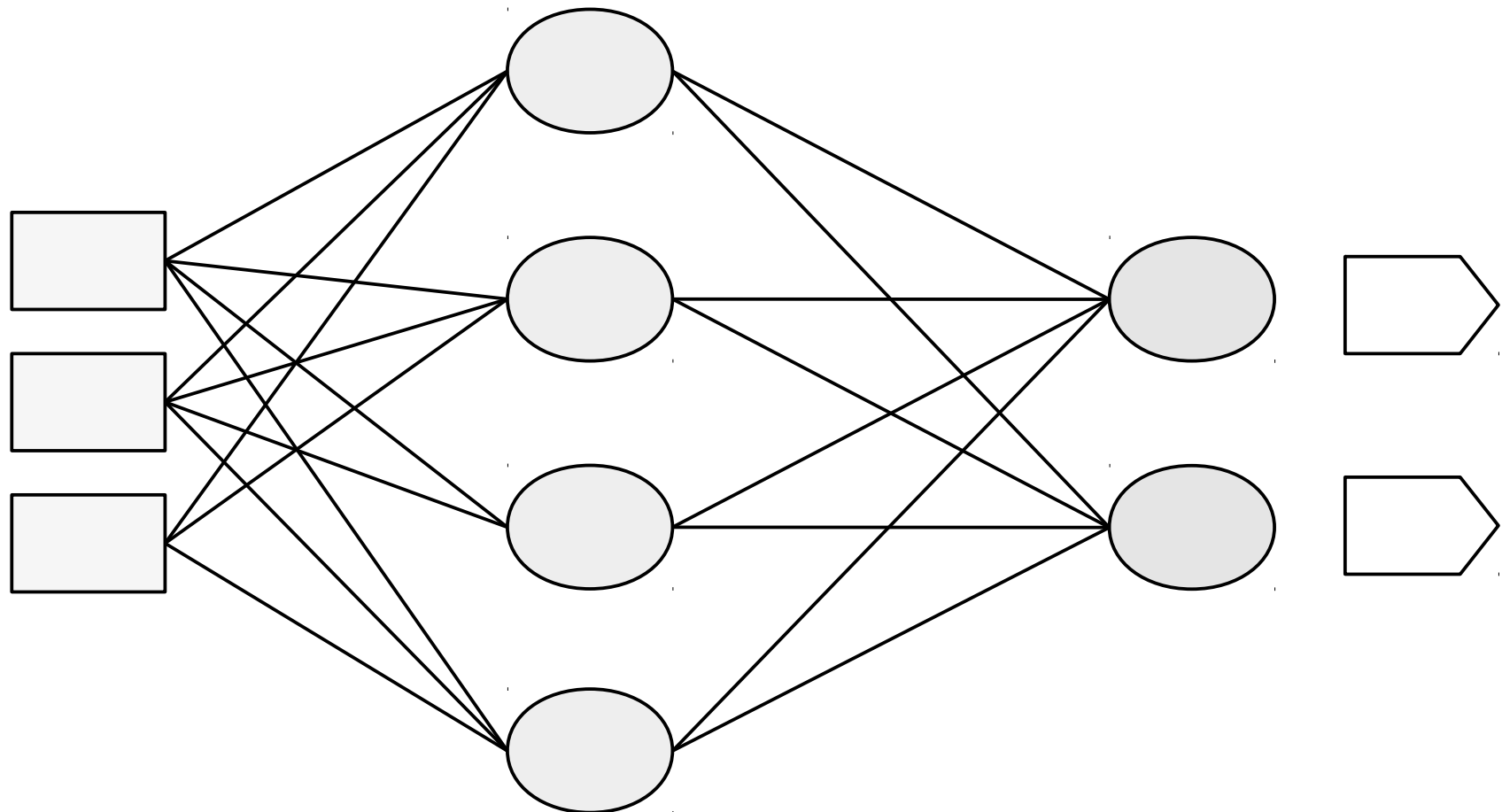
- Plano de Negócio
- Avaliação Econômica
- **Previsão de Mercado**
- Estimativas Iniciais de Custos

FEL-1

Decisões nos Portões



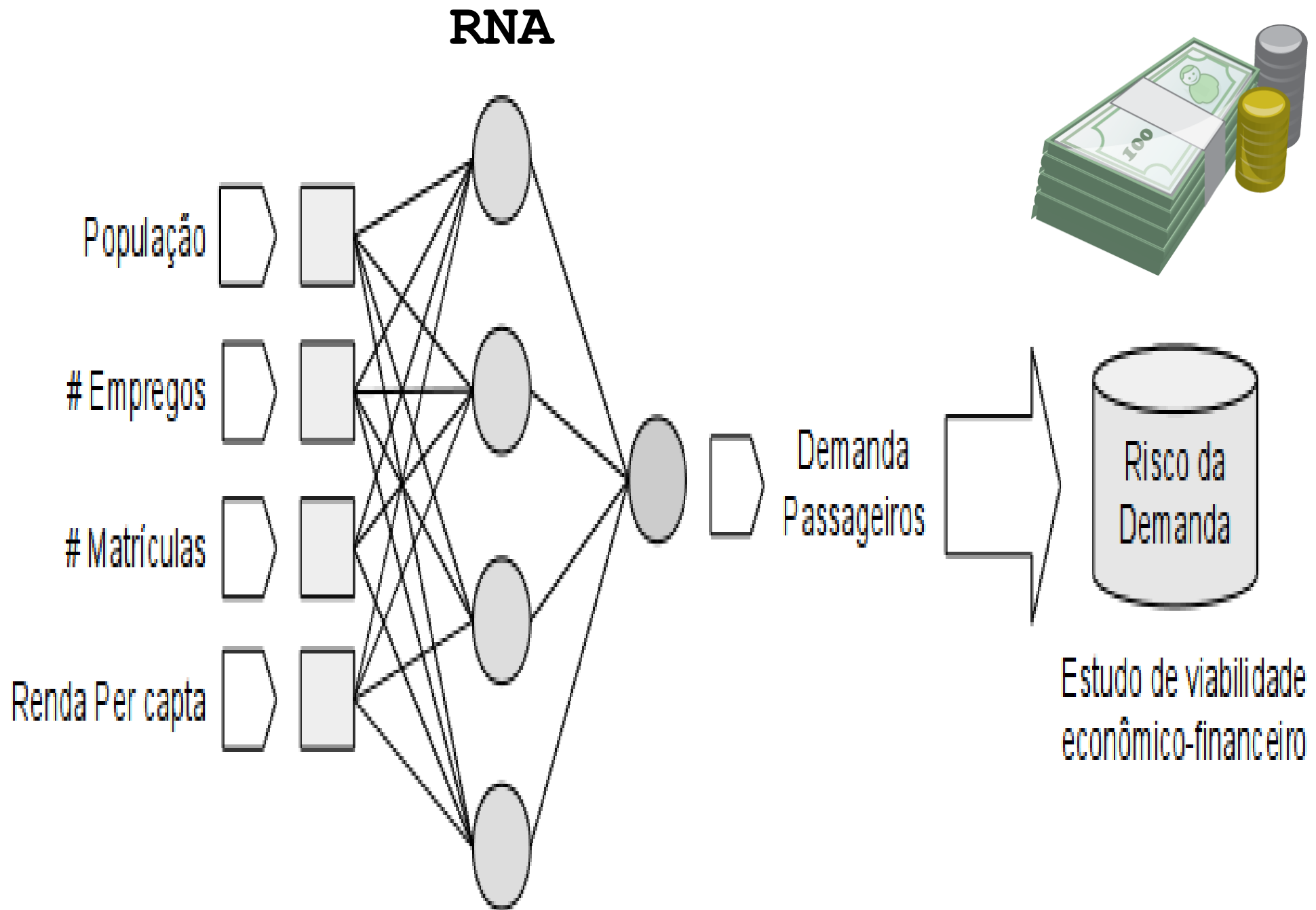
Redes Neurais Artificiais (RNA)

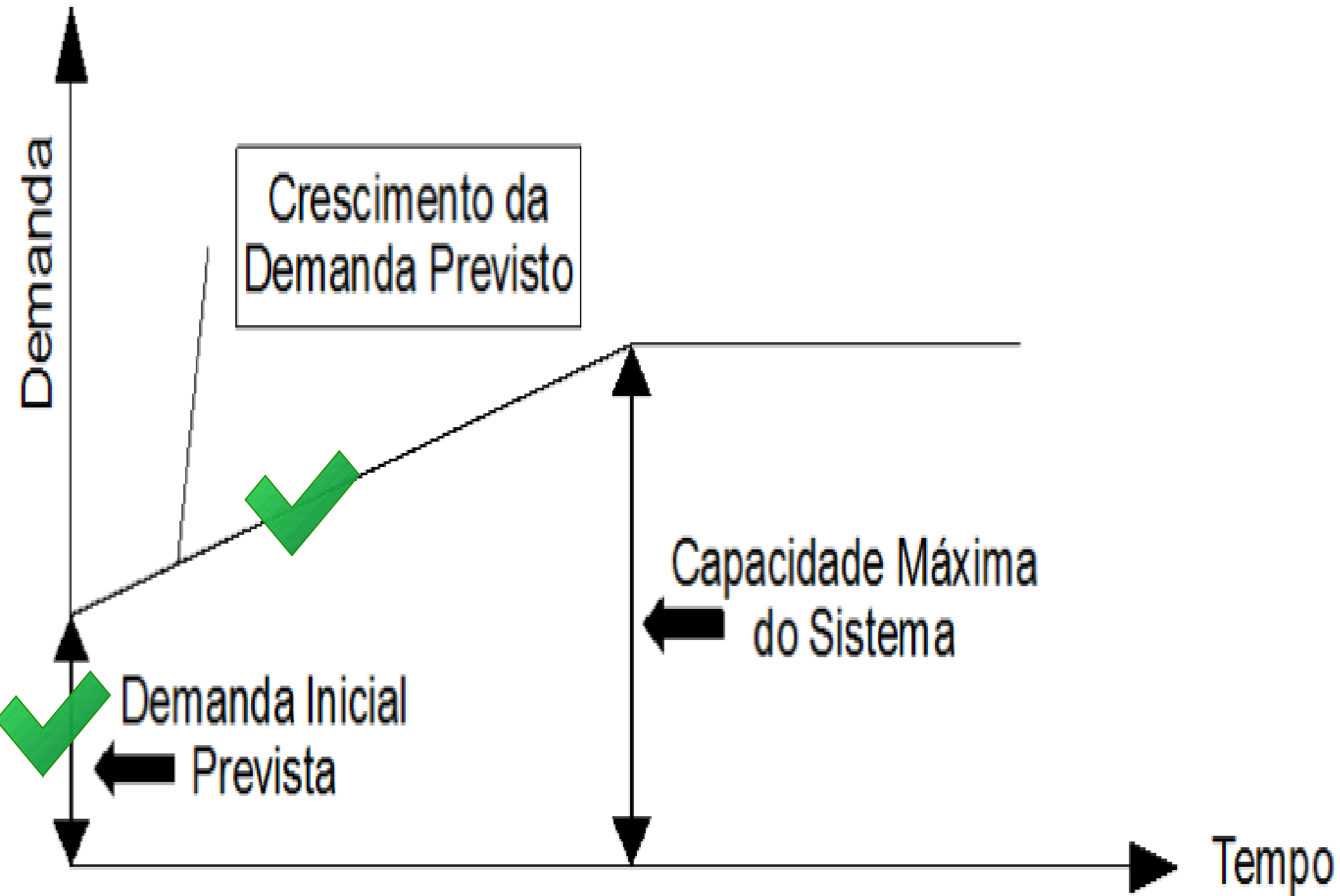


Dissertação

Previsão de demanda na fase de planejamento antecipado de projetos de transporte de passageiros: uma abordagem por redes neurais artificiais (2015)







Delineamento

Seleção das variáveis de entrada

Pré-processamento
Dados

Normalização

Reagrupamento

Configuração das RNAs

Treinamento, Teste e Validação

Métricas de desempenho

Seleção das variáveis de entrada 17

Realizada desde 1967 na RMSP

A cada dez anos (50 anos em 2017)

Objetivo: levantar informações sobre as viagens realizadas

Base para os estudos de planejamento de transporte

Em 2007 { 460 Zonas Pesquisa → 517
30 mil domicílios → 32

Análise Exploratória Pesq. O/D 2007
(Base aberta)

Pesq. O/D RMSP 50 anos

2017



Pesquisa O/D

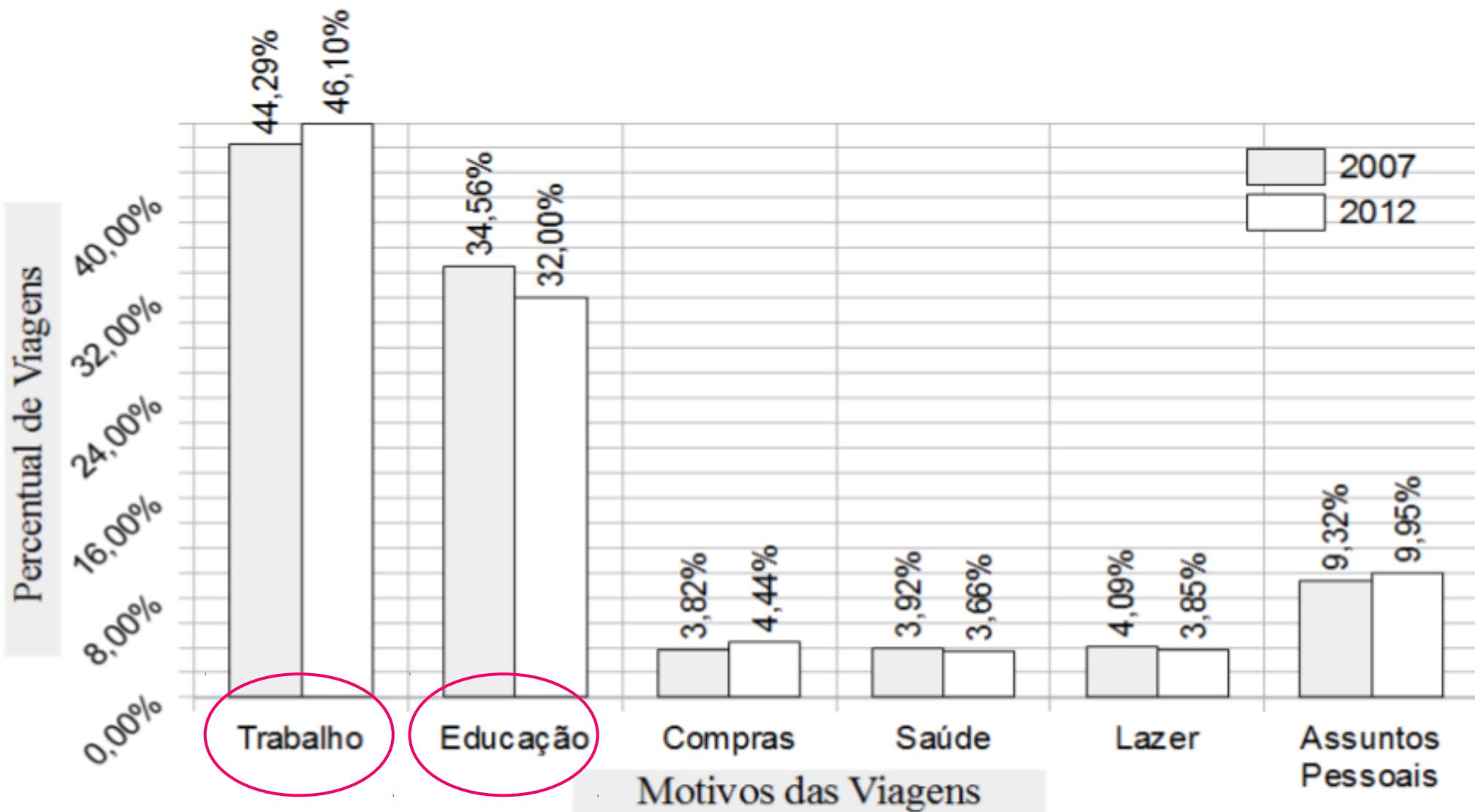


Figura 31: Percentual de viagens diárias na RMSP em função dos motivos nos anos de 2007 e 2012.

Fonte: Adaptado pelo autor de Metrô (2013b).

Aspecto Cultural

Tabela 1: Percentual de viagens/dia por modo e renda.

Renda [R\$] (Base out/2012)	Uso Transporte Coletivo	Uso Transporte Individual
Até 1.244	74,8%	25,2%
1.244 a 2.488	69,6%	30,4%
2.488 a 4.976	51,1%	48,9%
4.976 a 9.330	34,2%	65,8%
Mais de 9.330	24,1%	75,9%

Fonte: Adaptado pelo autor de Metrô (2013b).

**Variáveis
Independentes**

População

Empregos

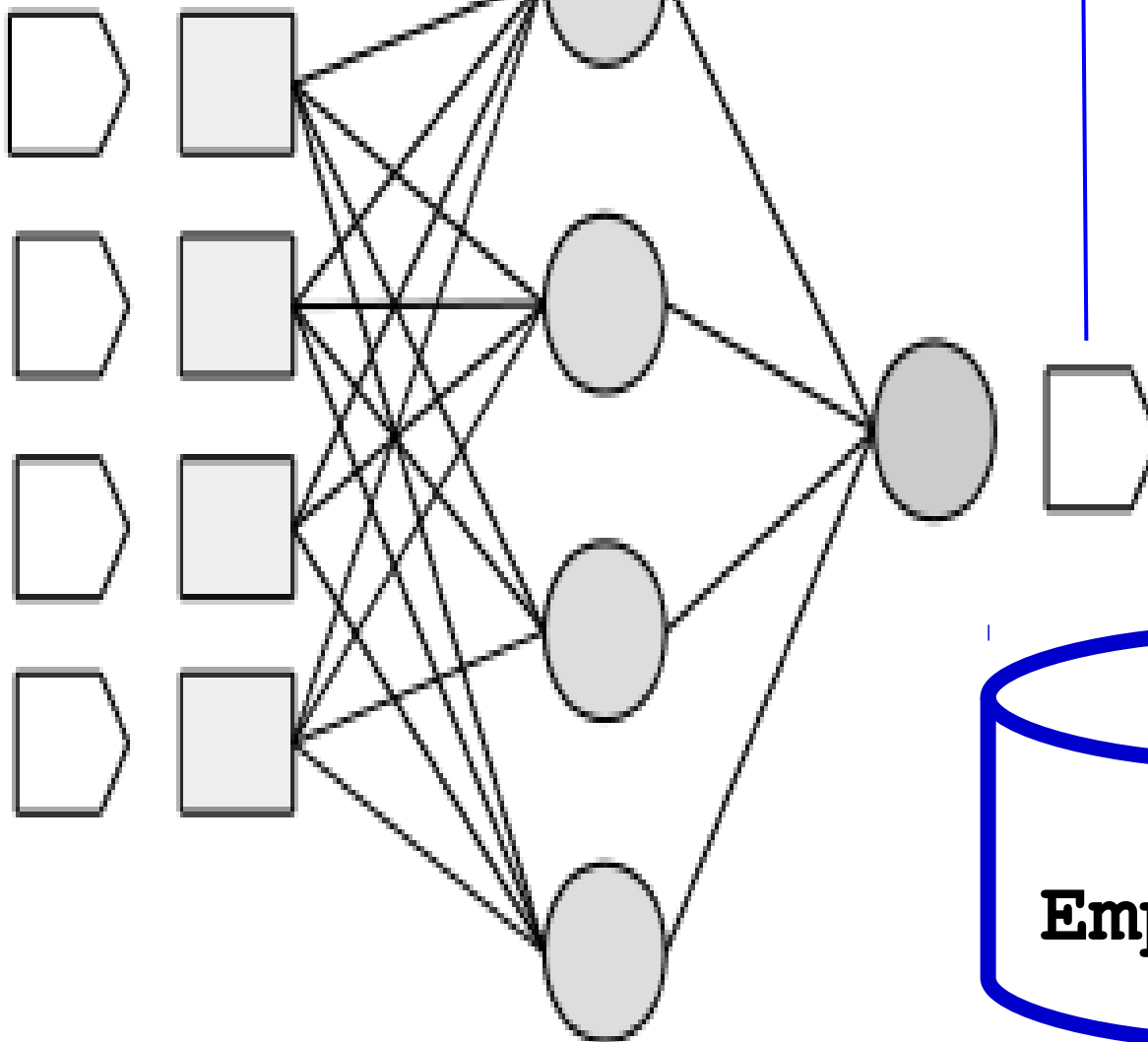
Matrículas

Renda Per capita

**Variáveis
Dependente**

Demanda
Passageiros

**BD
Empresas**



Realizada desde 1967 a cada 10 anos

1967

1977

1987

1997

2007

2017

2012

Pesquisa de Aferição

Pesq. Mobilidade

Urbana

31 zonas → RMSP

8,1 mil domicílios

Ideia inicial, só usar dados do Metrô-SP

Devido essa empresa atender só a cidade de São Paulo, algumas zonas possuíam várias estações, assim, a quantidade de dados era muito pequena.

1

São Paulo – Bela Vista, Bom Retiro, Brás, Cambuci, Consolação, Liberdade, Pari, República, Santa Cecília e Sé

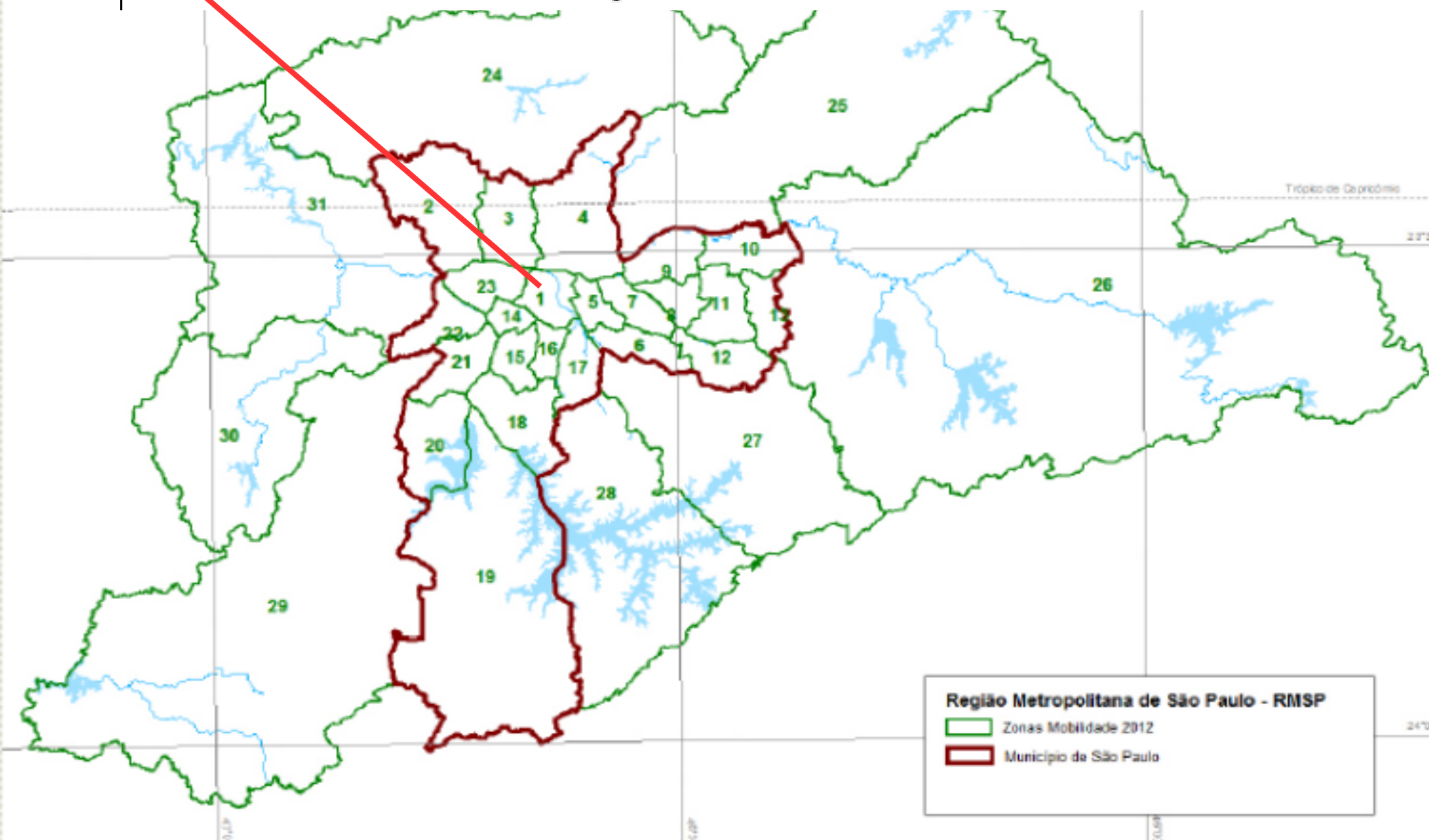


Figura 30: 31 zonas de tráfego da Pesquisa de Mobilidade 2012.

Fonte: Metrô (2013b).

Utilizar dados dos outros Operadores Grupo de Investigação

Operadoras	Total de Estações	Linhas	Quantidade de Estações
Metrô-SP	61 Estações	Linha 1 – Azul	23
		Linha 2 – Verde	14
		Linha 3 – Vermelha	18
		Linha 5 – Lilas	6
ViaQuatro	6 estações	Linha 4 – Amarela	6
CPTM	100 estações	Linha 7 – Rubi	18
		Linha 8 – Diamante	22
		Linha 9 – Esmeralda	18
		Linha 10 – Turquesa	13
		Linha 11 – Coral	16
		Linha 12 – Safira	13

Quadro 16: Linhas e Estações de metrô e trem metropolitano.

Fonte: Desenvolvido pelo autor.

Estações/Linhas por Zona de Tráfego

Linhas	Estações	
1	Vergueiro, São Joaquim, Liberdade, Sé, São Bento, Luz, Tiradentes e Armênia	
2	Brigadeiro ⁽¹⁶⁾ , Trianon Masp ⁽¹⁴⁾ , Clínicas ⁽¹⁴⁾ e consolação ⁽¹⁴⁾	
3	Deodoro, Santa Cecília, República, Anhangabaú, Sé, Pedro II, Brás ⁽⁵⁾ e Bresser ⁽⁵⁾	
4	Paulista, Luz e Republica	
7	Luz	
8	Júlio Prestes	
10	Brás ⁽⁵⁾ e Moóca ⁽¹⁾	
11	Luz e Brás ⁽⁵⁾	
12	Brás ⁽⁵⁾	
Total de Entrada de Passageiros		1.664.779

Quadro 18: Estações por linhas pertencentes à zona de tráfego 1

Pré-processamento dos Dados

Normalização

$$a = 0,1$$

$$b = 0,9$$

$$Valor_{normalizado} = (b - a) \frac{Valor - Valor_{Min}}{Valor_{Max} - Valor_{Min}} + a$$

(Lourencetti, 2011)



5.4.4

Computação Numérica

Open Source

Multiplataforma

ANN Toolbox 0.4.2.5



[Use cases](#)[Software](#)[Services](#)[About us](#)[Contact](#)

Data analytics: Engineers & Scientists need powerful numerical analysis

With over **100,000+ downloads per month**, Scilab is the **most open** numerical analysis and simulation software on the market.

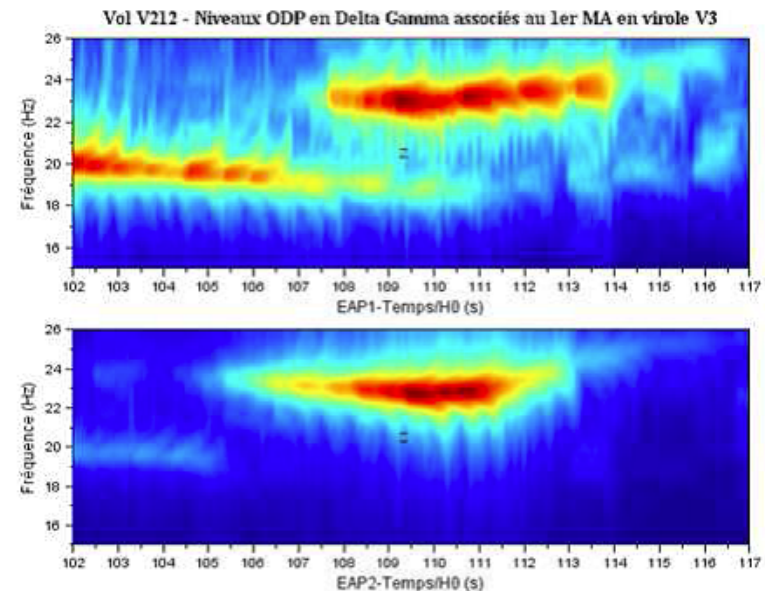
Flight Data Analysis

The application ALEx based on Scilab (stands for «Analyse des Lanceurs en Exploitation») was created by the French Space Agency CNES in Paris to analyze the measurement realized during the launch of rockets (particularly Ariane 5)

[READ THE USE CASE](#)

Neural Network Module

Livro



Dados Normalizados

Tabela 3: Dados normalizados das zonas de tráfego 1 até 15.

Zonas de tráfego													
1	2	3	4	5	6	7	8	9	10	11	13	14	15
0,267	0,336	0,485	0,127	0,290	0,203	0,192	0,260	0,397	0,225	0,268	0,100	0,145	
0,593	0,310	0,418	0,151	0,201	0,259	0,162	0,232	0,431	0,208	0,234	0,100	0,147	
0,900	0,185	0,385	0,184	0,145	0,217	0,100	0,163	0,187	0,125	0,107	0,349	0,440	
0,362	0,185	0,223	0,392	0,178	0,247	0,209	0,201	0,100	0,148	0,109	0,891	0,900	
0,900	0,116	0,200	0,149	0,133	0,170	0,152	0,130	0,120	0,163	0,124	0,302	0,100	

Fonte: Desenvolvido pelo autor

Pré-processamento dos Dados

Pré-processamento

**Reagrupamento
Dados**

Treinamento (50%)

Testes (25%)

Validação (25%)

Ordem randômica

Conj. Dados 1

Treinamento

Testes

Validação

3x

Conj. Dados 2

Treinamento

Testes

Validação

Conj. Dados 3

Treinamento

Testes

Validação

Pré-processamento dos Dados

Dados Reagrupados

Tabela 5: Grupos de dados de treinamento, teste e validação.

	Zonas de Tráfego		
	Conjunto de Dados 1	Conjunto de Dados 2	Conjunto de Dados 3
Treinamento	4, 11, 7, 10, 16, 15, 24, 5, 8, 31, 17 e 6	9, 14, 20, 24, 2, 21, 16, 5, 18, 8, 31 e 17	13, 24, 2, 6, 17, 9, 18, 26, 16, 7, 15 e 27
Teste	2, 13, 19, 18, 22 e 14	7, 1, 26, 13, 10 e 11	8, 23, 1, 10, 20 e 14
Validação	26, 1, 20, 27, 9, 23 e 21	4, 15, 22, 27, 23, 19 e 6	19, 31, 22, 4, 5, 21 e 11

Fonte: desenvolvido pelo autor.

Script: Geradados.sce

uma vez

► **DadosMSc.sod**



Camadas Escondidas

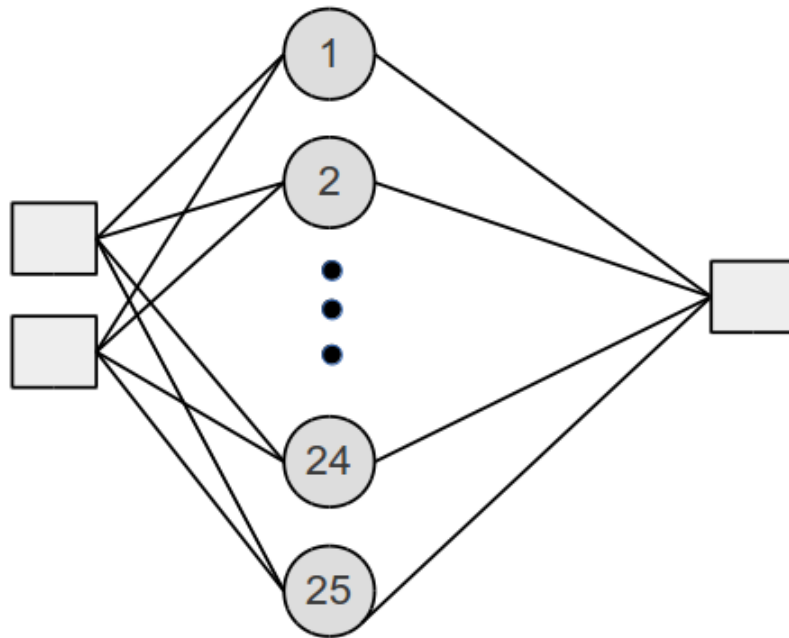
Qde neurônios por camada

Função de Ativação

Complexidade

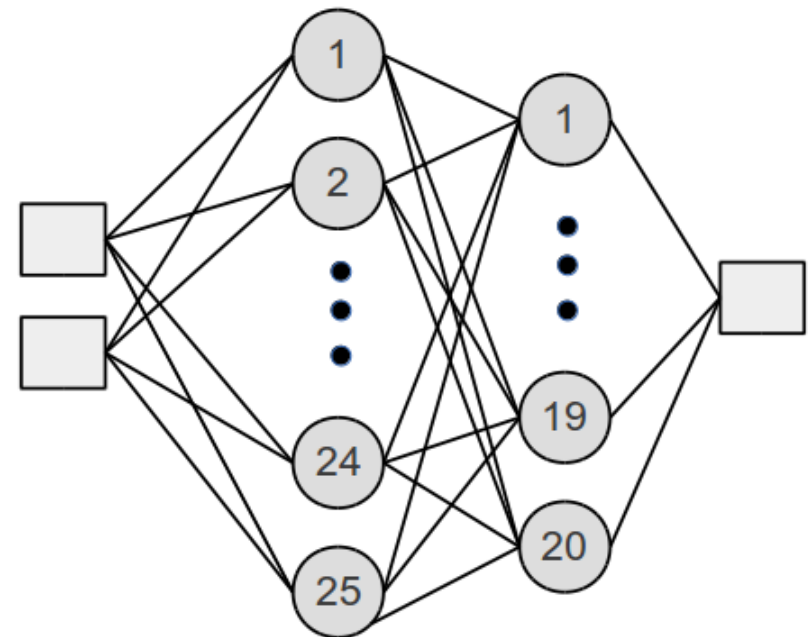
Empírica
(Nagai, 2006)

(Yao, 1997)



Qualquer função contínua

(Kovács, 2002)



Qualquer função matemática

Não há regras para determinação

Método de tentativa e erro

Heurísticas como ponto de partida
para 1ª camada oculta

$$q = \frac{E + S}{2}$$

$$q = 2.E + 1$$

$$q = \sqrt{E \cdot S}$$

Média das 3
heurísticas

Q = # nodos

E = # entradas da RNA

S = # saídas da RNA

Qde de Nodos

E = 4 entradas

S = 1 saída

População

Empregos

Matrículas

Renda per capta

Demanda de Passageiros

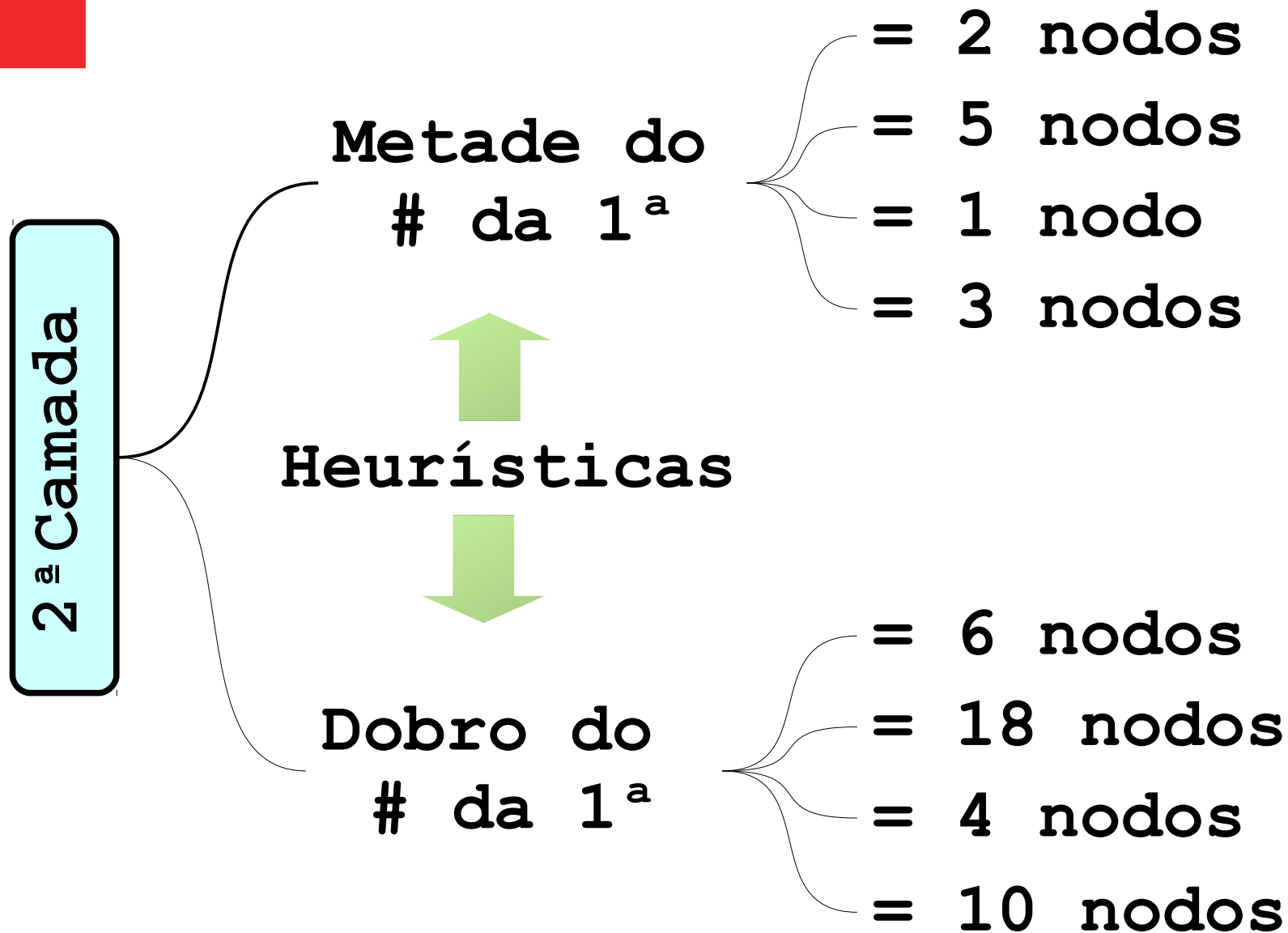
$$q = \frac{E + S}{2} = 2,5 \sim 3 \text{ nodos}$$

$$q = 2.E + 1 = 9 \text{ nodos}$$

$$q = \sqrt{E \cdot S} = 2 \text{ nodos}$$

Média das 3
heurísticas = 4,7 ~ 5 nodos

1ª Camada
oculta



Arq-01 → N=[4, 3, 1]

Arq-02 → N=[4, 9, 1]

Arq-03 → N=[4, 2, 1]

Arq-04 → N=[4, 5, 1]

Arq-05 → N=[4, 3, 2, 1]

Arq-06 → N=[4, 9, 5, 1]

Arq-07 → N=[4, 2, 1, 1]

Arq-08 → N=[4, 5, 3, 1]

Arq-09 → N=[4, 3, 6, 1]

Arq-10 → N=[4, 9, 18, 1]

Arq-11 → N=[4, 2, 4, 1]

Arq-12 → N=[4, 5, 10, 1]

4 entradas, 1
camada escondida
com 3 nodos, uma
saída

4 entradas, 2
camadas
escondidas, a 1^a
com 9 nodos, a 2^a
com 5 nodos, uma
camada de saída

Tipo de RNA:

Multilayer Perceptron (**MLP**)

Função de Ativação: **Sigmoidal**

Algoritmo treinamento:

Back-Propagation

Taxa de Aprendizagem

(suavidade de atualização dos pesos)

0,01

0,50

0,90

Termo Momento

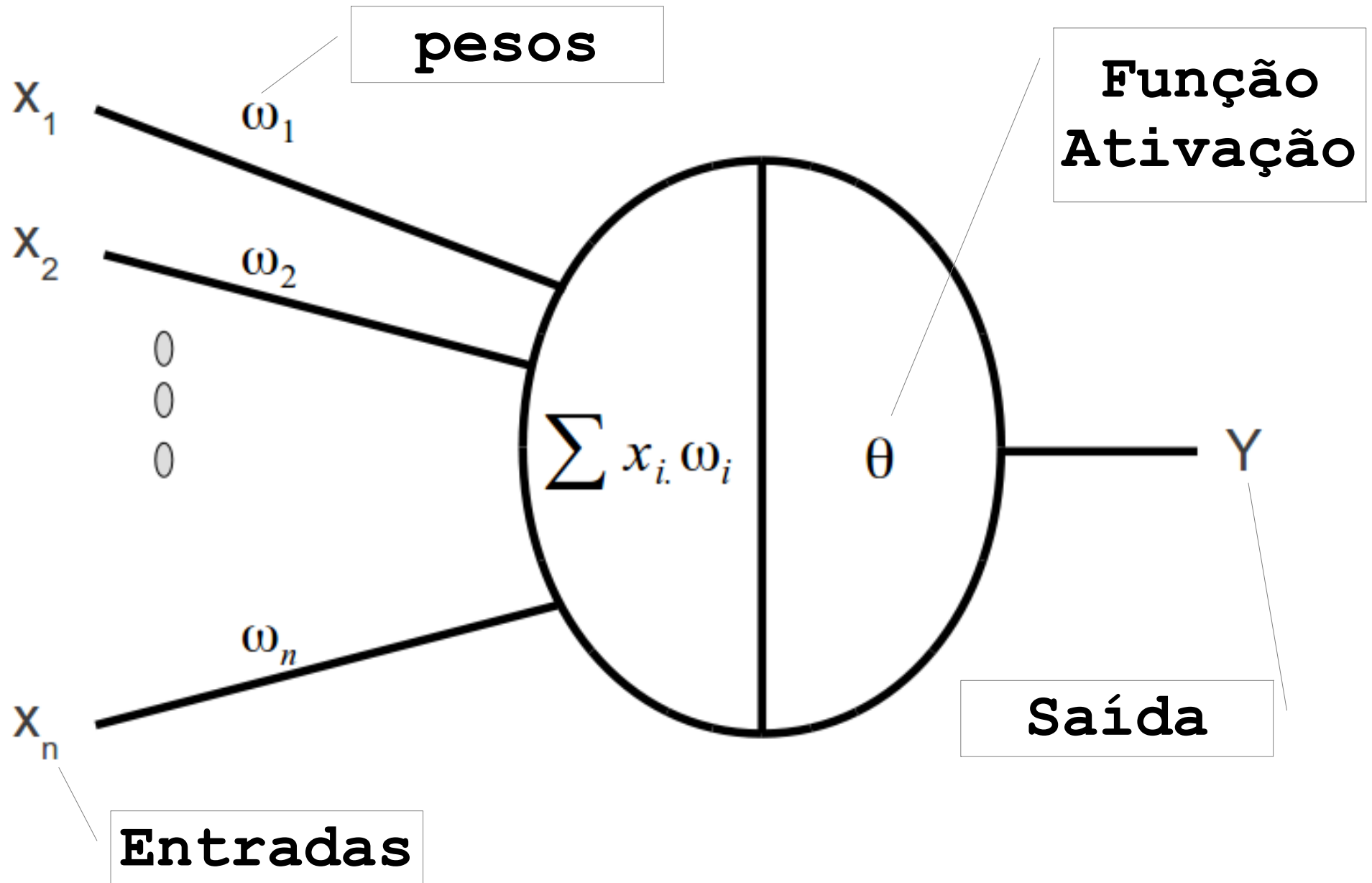
(Se necessário, acelera o treinamento)

0,50

0,90

(Raia, 2000)

Outras ...



Critérios
de Parada

Tolerância = 0,0001

épocas = 10.000

Desempenho

Erro Quadrático Médio

$$EQM = \frac{1}{N} \sum_{i=1}^N (Demanda_{Real} - Demanda_{Prevista})^2$$

Coeficientes

Correlação (r)

Determinação (R^2)

Relevância das variáveis
Dependentes → Garson

Scripts

Treinamento.sce

Teste.sce

Validacao.sce

Dispersao.sce

Treinamento, Teste e Validação

39

Tabela 29: Arq-12— Tolerância de 0,0001 e nº épocas de 10.000 iterações.

A	M	Desempenho	Etapas								
			Treinamento			Teste			Validação		
			Dados1	Dados2	Dados3	Dados1	Dados2	Dados3	Dados1	Dados2	Dados3
0,01	0,5	Tempo [s]	74,218	76,263	64,256	47,622	65,795	37,203	0,003	0,003	0,003
		Tempo [CPU]	74,496	76,672	64,112	50,380	66,892	37,744	0,008	0,016	0,008
		EQM [%]	0,124	0,248	0,148	0,403	0,360	0,560	7,958	4,008	2,831
0,5	0,5	Tempo [s]	61,998	62,790	62,937	35,628	35,649	36,416	0,003	0,005	0,001
		Tempo [CPU]	62,792	63,984	63,484	35,732	35,944	37,160	0,008	0,015	0,000
		EQM [%]	0,112	0,033	0,066	0,020	0,013	0,022	3,176	4,103	0,051
0,5	0,9	Tempo [s]	62,823	62,631	62,388	35,754	36,502	36,295	0,002	0,002	0,002
		Tempo [CPU]	63,044	62,724	62,452	35,864	37,520	36,620	0,008	0,012	0,008
		EQM [%]	0,037	0,031	0,042	0,010	0,014	0,024	0,640	3,834	0,058
0,9	0,5	Tempo [s]	62,553	62,658	63,538	36,305	36,040	36,793	0,002	0,001	0,003
		Tempo [CPU]	62,728	62,896	63,764	36,792	36,300	36,828	0,008	0,004	0,004
		EQM [%]	0,095	0,031	0,060	0,012	0,011	0,022	1,993	2,872	0,050
0,9	0,9	Tempo [s]	62,640	62,378	62,231	34,911	36,074	35,655	0,004	0,003	0,002
		Tempo CPU	62,704	62,636	62,360	35,484	36,576	35,836	0,012	0,012	0,004
		EOM[%]	0.019	0.028	0.005	0.002	0.013	0.017	0.624	3.310	0.045

Tabela 30: Valores previstos e coeficientes de correlação e determinação da reta de 45° da melhor Arq-12.

Valores Previstos	0,137	0,173	0,173	0,204	0,182	0,171	0,158
Coeficiente de Correlação (r)	0,478			Coeficiente de Determinação (R^2)			0,228

Fonte: saídas do SciLab, adaptado pelo autor.

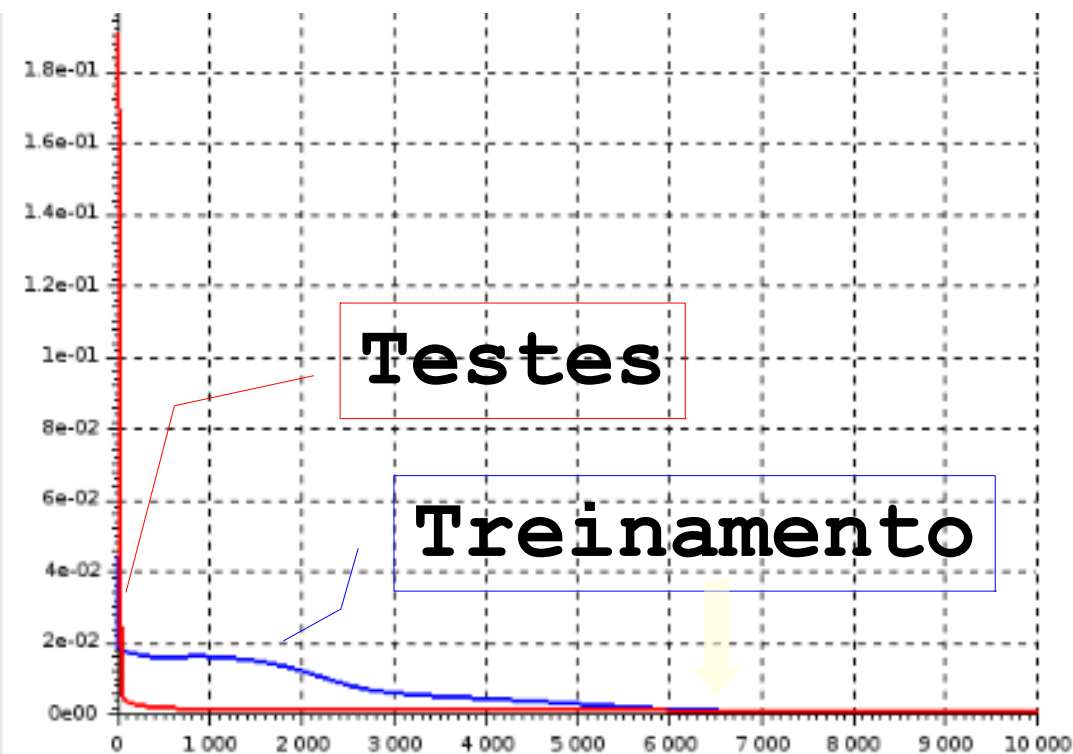


Figura 55: Erro x Época da melhor Arq-12.

Fonte: saída do SciLab.

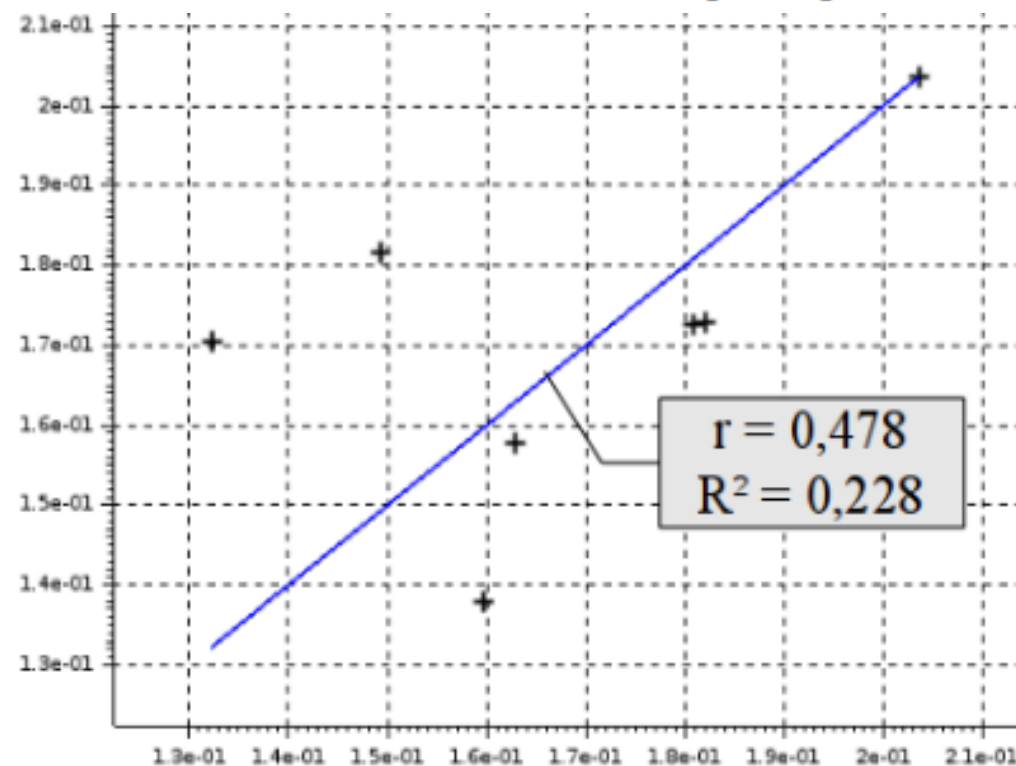


Figura 56: Gráfico de dispersão - valores reais x previstos da melhor Arq-12.

Tabela 31: Resumo dos melhores resultados obtidos ordenados pelo EQM%

Arquitetura		Taxa de Aprendizagem/ Momento	Conjunto de Dados	Tempos [s]		EQM% Validação	Coeficientes	
Nº	Topologia			Treina- mento	Teste		\hat{r}	R^2
Arq-12	[4, 5, 10, 1]	0,9/0,9	Dados 3	62,231	35,655	0,045	0,478	0,228
Arq-06	[4, 9, 5, 1]	0,5/0,5	Dados 3	63,983	36,438	0,050	0,288	0,083
Arq-05	[4, 3, 2, 1]	0,9/0,5	Dados 3	62,761	36,673	0,053	0,209	0,043
Arq-09	[4, 3, 6, 1]	0,5/0,9	Dados 3	63,387	36,257	0,053	0,480	0,230
Arq-07	[4, 2, 1, 1]	0,9/0,5	Dados 3	61,079	35,006	0,057	0,145	0,021
Arq-11	[4, 2, 4, 1]	0,9/0,5	Dados 3	63,052	36,037	0,061	0,031	0,001
Arq-08	[4, 5, 3, 1]	0,5/0,5	Dados 3	143,882	35,634	0,064	0,419	0,175
Arq-03	[4, 2, 1]	0,9/0,9	Dados 3	48,511	28,976	0,078	0,290	0,084
Arq-10	[4, 9, 18, 1]	0,5/0,5	Dados 3	67,105	38,176	0,109	0,491	0,241
Arq-02	[4, 9, 1]	0,5/0,9	Dados 3	54,290	38,148	0,138	0,605	0,366
Arq-01	[4, 3, 1]	0,5/0,9	Dados 3	57,853	39,561	0,189	0,459	0,210
Arq-04	[4, 5, 1]	0,5/0,5	Dados 3	49,475	29,454	0,543	0,052	0,003

Fonte: saídas do SciLab, adaptado pelo autor

Duas camadas Ocultas

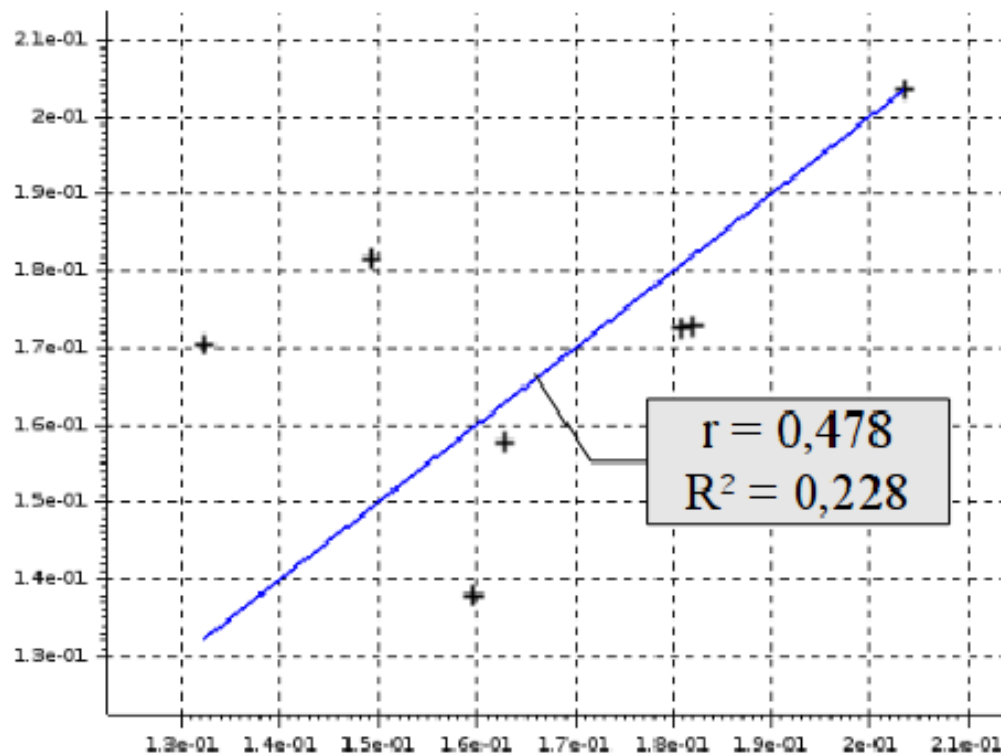


Figura 56: Gráfico de dispersão - valores reais x previstos da melhor Arq-12.

Arq-12: $N=[4,5,10,1]$

EQM% = 0,045

Uma camadas Ocultas

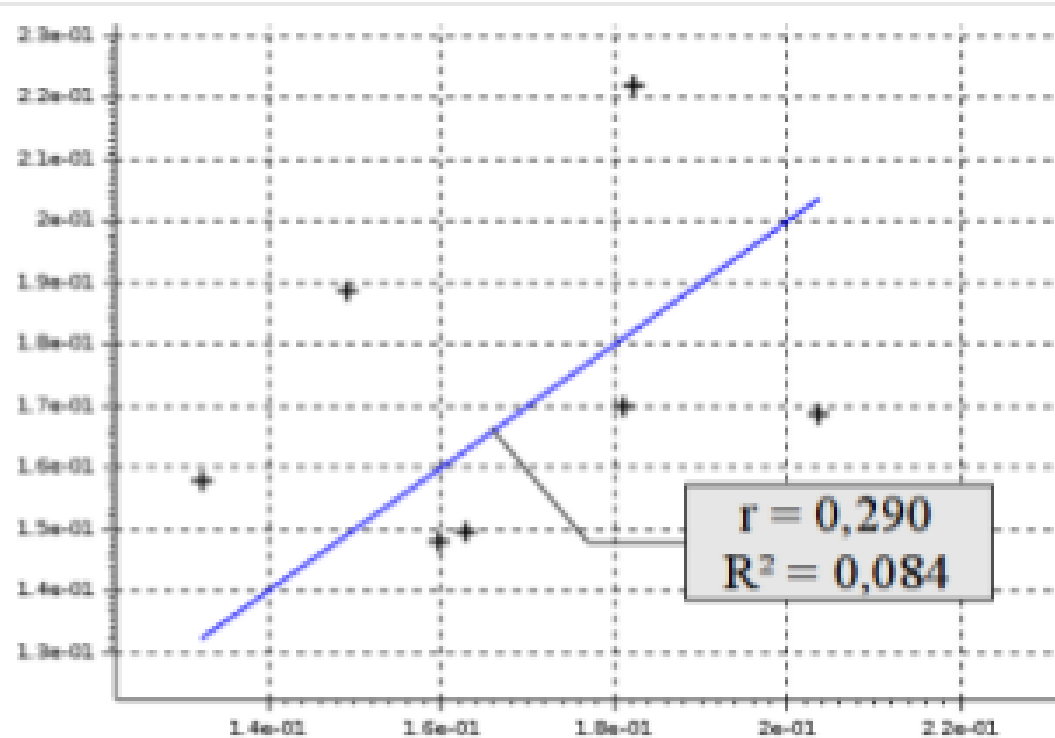


Figura 38: Gráfico de dispersão - valores reais x previstos da melhor Arq-03.

Arq-03: $N=[4,2,1]$

EQM% = 0,078

Método de Garson - Arq-03

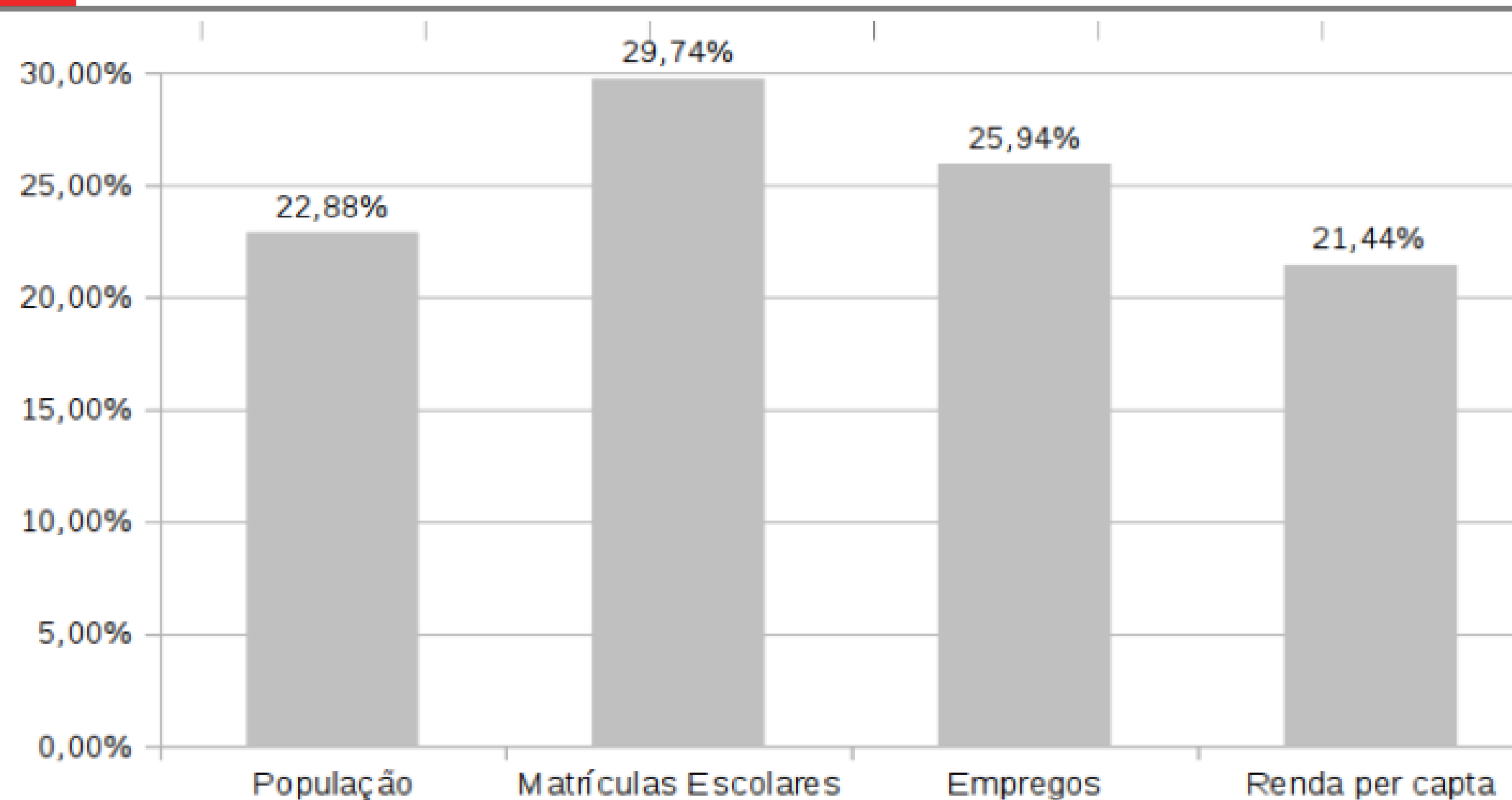


Figura 58: Relevâncias das variáveis de entradas, segundo método de Garson.

Média = 25,00% e desvio padrão = 3,68%

Fonte: desenvolvido pelo autor.

Arq-03 N=[4,2,1]

Tabela 37: Resumo dos resultados obtidos para as novas arquiteturas.

Arquitetura		Taxa de Aprendizagem/ Momento	Conjunto De Dados	Tempos [s]		EQM% Validação
Nº	N			Treinamento	Teste	
Arq-031	[3, 2, 1]	0,9/0,9	Dados 3	48,408	28,621	0,086
Arq-032	[2, 2, 1]	0,9/0,9	Dados 3	50,745	29,766	3,662

Fonte: saídas do SciLab, adaptado pelo autor.



+10% Arq-03

+4595% Arq-03

Aplicações Avançadas em Data Science

Previsão de Demanda de Passageiros em
Sistemas Metroferroviários:

uma abordagem por **Redes Neurais Artificiais**

Vagner Sanches Vasconcelos

vsvasconcelos@gmail.com

 **GitHub** /vsvasconcelos/PyData-SP_17052017