

6 Aplicação da Metodologia Proposta

6.1 Introdução

A utilização da metodologia Box & Jenkins para a identificação da estrutura adequada é tarefa difícil para um usuário sem conhecimento prévio. Por isso, sistemas especialistas têm sido utilizada nesta identificação.

Para as três metodologias aplicadas – árvores de decisão, redes neurais e KNN –, as variáveis de entrada, que são os vetores formados pelos estimadores da autocorrelação e da autocorrelação parcial, são determinados da seguinte forma:

- Modelos Box & Jenkins não sazonais

Utilizam-se os estimadores das autocorrelação e autocorrelação parcial nos *lags* 1, 2, 3, 4 e 5. Estes estimadores são gerados através de simulações de séries dos modelos AR(1), AR(2), MA(1), MA(2) e ARMA(1,1) e Ruído branco com 300, 350, 400, 450, 500 e 600 observações para cada um deles. Para cada série gerada, forma-se o vetor das variáveis de entrada $[ACF1 \ ACF2 \ ACF3 \ ACF4 \ ACF5 \ PACF1 \ PACF2 \ PACF3 \ PACF4 \ PACF5]$, denominado vetor dos estimadores. Ao todo tem-se 3600 vetores de entrada para o conjunto de treinamento e 864 vetores para o conjunto de teste.

- Modelos Box & Jenkins sazonais de período 3

Utilizam-se os estimadores das autocorrelação e autocorrelação parcial nos *lags* 3, 6, 9, 12 e 15. Estes estimadores são gerados através de simulações de séries dos modelos SAR(1), SAR(2), SMA(1), SMA(2) e SARMA(1,1) e Ruído branco com 300, 350, 400, 450, 500 e 600 observações para cada um deles. Para cada série gerada, forma-se o vetor das variáveis de entrada $[ACF3 \ ACF6 \ ACF9 \ ACF12 \ ACF15 \ PACF3 \ PACF6 \ PACF9 \ PACF12 \ PACF15]$,

denominado vetor dos estimadores. Ao todo tem-se 3600 vetores de entrada para o conjunto de treinamento e 864 vetores para o conjunto de teste.

- Modelos Box & Jenkins sazonais de período 4

Utilizam-se os estimadores das autocorrelação e autocorrelação parcial nos *lags* 4, 8, 12, 16 e 20. Estes estimadores são gerados através de simulações de séries dos modelos SAR(1), SAR(2), SMA(1), SMA(2) e SARMA(1,1) e Ruído branco com 300, 350, 400, 450, 500 e 600 observações para cada um deles. Para cada série gerada, forma-se o vetor das variáveis de entrada $[ACF4 \ ACF8 \ ACF12 \ ACF16 \ ACF20 \ PACF4 \ PACF8 \ PACF12 \ PACF16 \ PACF20]$, denominado vetor dos estimadores. Ao todo tem-se 3600 vetores de entrada para o conjunto de treinamento e 864 vetores para o conjunto de teste.

- Modelos Box & Jenkins sazonais de período 6

Utilizam-se os estimadores das autocorrelação e autocorrelação parcial nos *lags* 6, 12, 18, 24 e 30. Estes estimadores são gerados através de simulações de séries dos modelos SAR(1), SAR(2), SMA(1), SMA(2) e SARMA(1,1) e Ruído branco com 300, 350, 400, 450, 500 e 600 observações para cada um deles. Para cada série gerada, forma-se o vetor das variáveis de entrada $[ACF6 \ ACF12 \ ACF18 \ ACF24 \ ACF30 \ PACF6 \ PACF12 \ PACF18 \ PACF24 \ PACF30]$, denominado vetor dos estimadores. Ao todo tem-se 3600 vetores de entrada para o conjunto de treinamento e 864 vetores para o conjunto de teste.

- Modelos Box & Jenkins sazonais de período 12

Utilizam-se os estimadores das autocorrelação e autocorrelação parcial nos *lags* 12, 24, 36, 48 e 60. Estes estimadores são gerados através de simulações de séries dos modelos SAR(1), SAR(2), SMA(1), SMA(2) e SARMA(1,1) e Ruído branco com 300, 350, 400, 450, 500 e 600 observações para cada um deles. Para cada série gerada, forma-se o vetor das variáveis de entrada $[ACF12 \ ACF24 \ ACF36 \ ACF48 \ ACF60 \ PACF12 \ PACF24 \ PACF36 \ PACF48 \ PACF60]$,

denominado vetor dos estimadores. Ao todo tem-se 3600 vetores de entrada para o conjunto de treinamento e 864 vetores para o conjunto de teste.

6.2 Metodologia aplicada à árvore de decisão

A partir dos modelos acima descritos, passa-se a traçar a metodologia aplicada à árvore de decisão.

Constrói-se uma árvore de decisão que classifique cada modelo Box & Jenkins levando em consideração o comportamento teórico da autocorrelação e da autocorrelação parcial. As entradas serão os vetores estimadores e as saídas serão os modelos SAR(1), SAR(2), SMA(1), SMA(2), SARMA(1,1) e Ruído branco no caso dos modelos sazonais e AR(1), AR(2), MA(1), MA(2), ARMA(1,1) e Ruído branco no caso dos modelos não sazonais.

Para a construção da árvore é utilizado o CART. As características da árvore de classificação são:

- Critério de divisão: índice de Gini
- Cálculo do custo da árvore: método da ressubstituição
- Podagem: árvore que possui custo mínimo
- Entrada: autocorrelação e autocorrelação parcial
- Saída: identificação do modelo

6.3 Metodologia aplicada às redes neurais

Já em relação à metodologia aplicada às redes neurais, estuda-se qual será a melhor arquitetura de redes neurais que classifique cada modelo Box & Jenkins levando em consideração o comportamento teórico da autocorrelação e da autocorrelação parcial. Como será uma rede neural para fazer classificação com um algoritmo de aprendizado supervisionado, a arquitetura de rede escolhida foi a MLP (*Multi Layer Perceptron*).

Para esta implementação será utilizado apenas um dos modelos Box & Jenkins, pois todos os modelos possuem o mesmo tipo de conjunto de treinamento e teste. As entradas serão os vetores estimadores e as saídas serão os modelos SAR(1), SAR(2), SMA(1), SMA(2), SARMA(1,1) e Ruído branco no caso dos

modelos sazonais e AR(1), AR(2), MA(1), MA(2), ARMA(1,1) e Ruído branco no caso dos modelos não sazonais.

Para a classificação dos modelos, foram simuladas diversas arquiteturas com diversos números de neurônios, diferentes funções de propagação e com uma e duas camadas escondidas. Após várias simulações, a arquitetura encontrada que maximizou o percentual de acerto no conjunto de teste foi uma rede com três camadas, sendo duas delas escondidas. A primeira camada escondida contém 31 neurônios, a segunda contém 10 neurônios e a camada de saída possui 6 neurônios. Nas duas primeiras camadas a função de propagação a ser utilizada é a tangente hiperbólica sigmóide e na terceira camada, a função logística sigmóide. O treinamento é feito por batelada, isto é, o ajuste dos pesos é feito ao final de cada época, e o algoritmo de treinamento é o RPROP (*Resilient backpropagation algorithm*). Esta arquitetura será a mesma para todos os modelos Box & Jenkins sazonais ou não sazonais. A figura abaixo representa a arquitetura da rede neural utilizada neste trabalho.

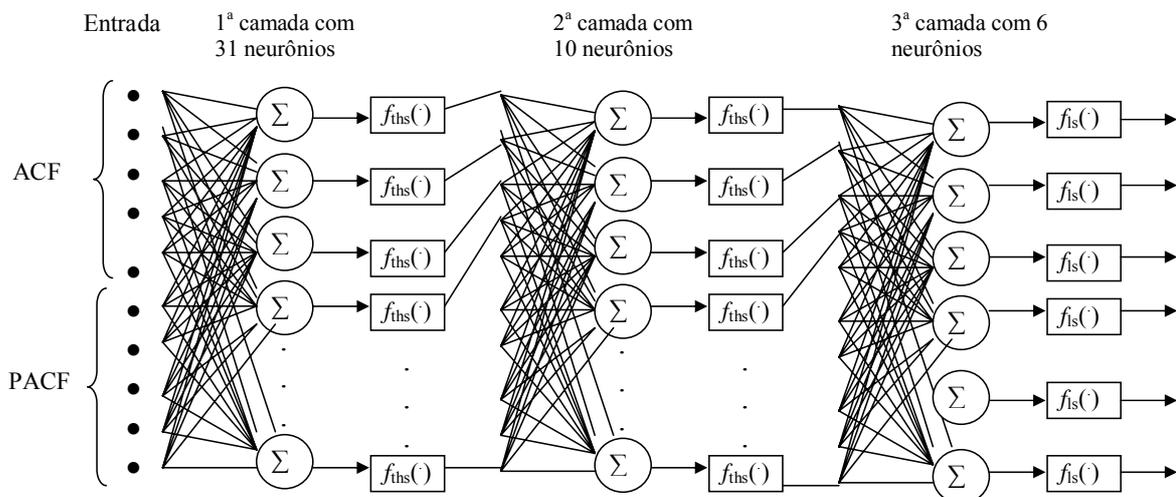


Figura 20 – Representação da arquitetura da rede neural

6.4

Metodologia aplicada a KNN (*K-Nearest Neighbors*)

Esta ferramenta tem a finalidade de determinar a classe de um objeto por meio de seus K vizinhos mais próximos sendo assim, esse objeto terá a mesma classe que a maioria de seus vizinhos mais próximo. As entradas serão os vetores estimadores e as saídas serão os modelos SAR(1), SAR(2), SMA(1), SMA(2), SARMA(1,1) e Ruído branco no caso dos modelos sazonais e AR(1), AR(2), MA(1), MA(2), ARMA(1,1) e Ruído branco no caso dos modelos não sazonais.

6.5

Resultados obtidos

Para comparar o percentual de acerto das árvores de decisão, das redes neurais e do KNN, é utilizado o *software Forecast Pro for Windows* Versão 3.50, que permite identificar os modelos Box & Jenkins em todos os conjuntos de teste propostos neste trabalho.

Os conjuntos utilizados para treinamento e teste são os mesmos para os quatro classificadores.

Quando os parâmetros dos modelos Box & Jenkins – definidos no capítulo 2 – estão próximos de zero, as séries geradas tornam-se ruído branco. Devido a este fato, o erro na classificação torna-se mais alto. Então, com o intuito de reduzir esse erro, foram expurgados todos os vetores de estimadores, no conjunto de treinamento e teste, que representavam incorretamente um modelo. Para facilitar a compreensão, a partir deste momento este conjunto será chamado de conjunto dos vetores de estimadores modificados.

Para demonstrar esse alto índice de erro, foram criados dois novos conjuntos de estimadores para os modelos sazonais de período 12 com as seguintes características: valores de $-0,2 < \Theta < 0,2$ e $-0,2 < \Phi < 0,2$ e $|\Theta| > 0,2$ e $|\Phi| > 0,2$. O conjunto de treinamento e o conjunto de teste são formados por 600 vetores.

Box & Jenkins define o ruído branco a partir de uma distribuição Normal com média zero e variância igual a 1 (capítulo 2). Neste trabalho será feito uma simulação de séries sazonais de período 12 tendo o ruído branco a partir de uma

distribuição Normal com média zero e variância igual a 2 para verificar qual será a alteração no percentual do erro.

Nos modelos não sazonais e sazonais de período 12 serão também, verificados os percentuais de acerto/erro para o subconjunto do conjunto de vetores de estimadores e o de estimadores modificados formados pelos lags 1, 2, 3 e 4 (modelos não sazonais) e 12, 24, 36 e 48 (modelos sazonais).

Os resultados obtidos com todos estes conjuntos modificados encontram-se no apêndice.

Neste trabalho será realizada a classificação com o conjunto de estimadores inicial, sem alterações, em todos os classificadores.

6.5.1 Resultados obtidos ao utilizar **Árvore de Decisão**

6.5.1.1 Análise dos resultados para os modelos **Box & Jenkins não-sazonais**.

- o Em relação ao conjunto de treinamento e teste inicial.

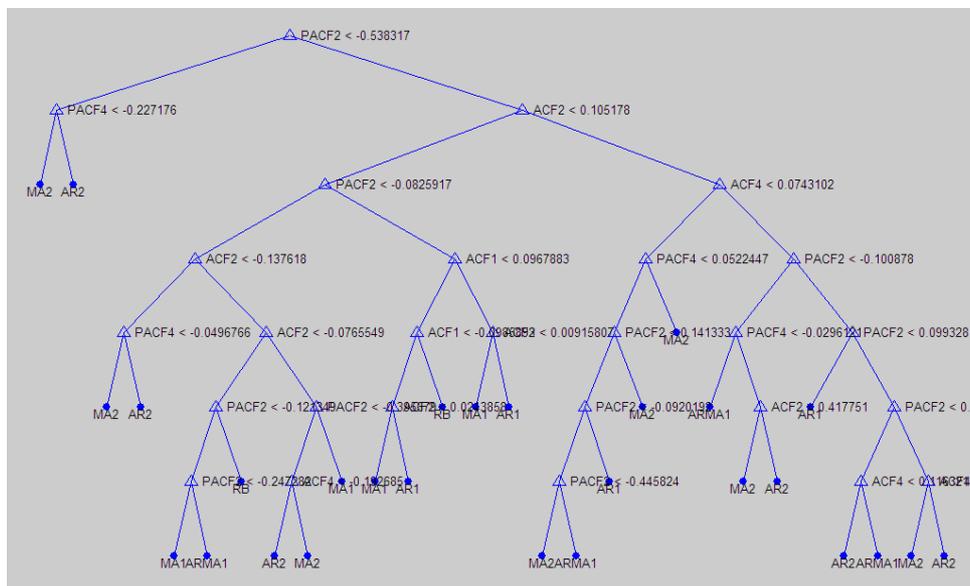


Figura 21 - **Árvore de decisão para modelos não-sazonais**

Para este conjunto de treinamento e teste, o erro no conjunto de teste é igual a 30.44% e o número ótimo de nós é de 55, sendo que 28 são nós terminais.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Árvore Modelo Não-sazonal

Performance Percentual no Treinamento

	% de Acerto	% de Erro
AR1	90	10
AR2	90.3333	9.6667
MA1	88.6667	11.3333
MA2	90.1667	9.8333
ARMA1	76.3333	23.6667
RB	94.3333	5.6667
Média	88.3056	11.6944

Performance Percentual no Teste

	% de Acerto	% de Erro
AR1	79.1667	20.8333
AR2	61.1111	38.8889
MA1	75.6944	24.3056
MA2	67.3611	32.6389
ARMA1	42.3611	57.6389
RB	91.6667	8.3333
Média	69.5602	30.4398

6.5.1.2

Análise dos resultados para os modelos Box & Jenkins sazonais

➤ **Em relação à sazonalidade de período 3**

- Em relação ao conjunto de treinamento e teste inicial.

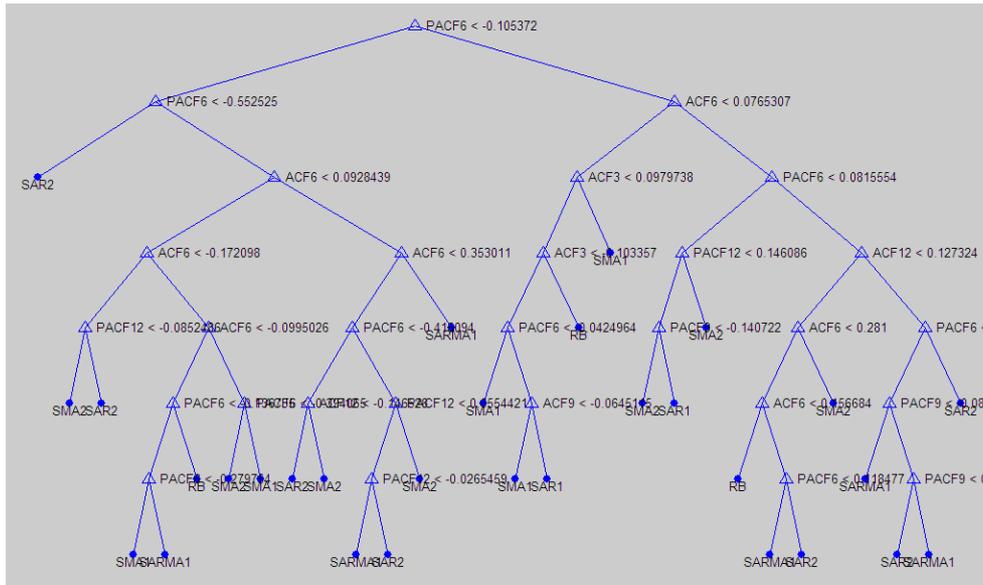


Figura 22 - Árvore de decisão para modelos sazonais de período 3

Para este conjunto de treinamento e teste, o erro no conjunto de teste é igual a 29.98% e o número ótimo de nós é de 59, sendo que 30 são nós terminais.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Para este conjunto de treinamento e teste, o erro no conjunto de teste é igual a 31.83% e o número ótimo de nós é de 97, sendo que 47 são nós terminais. Esta árvore possui mais complexidade que a árvore dos modelos sazonais de período 3 pois possui um número maior de nós.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Árvore Modelo Sazonal Período 4

Performance Percentual no Treinamento

	% de Acerto	% de Erro
SAR1	87	13
SAR2	91	9
SMA1	87.5	12.5
SMA2	87	13
SARMA1	74.8333	25.1667
RB	94.1667	5.8333
Média	86.9167	13.0833

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	71.5278	28.4722
SAR2	66.6667	33.3333
SMA1	73.6111	26.3889
SMA2	67.3611	32.6389
SARMA1	36.1111	63.8889
RB	93.75	6.25
Média	68.1713	31.8287

➤ **Em relação à sazonalidade de período 6**

- Em relação ao conjunto de treinamento e teste inicial.

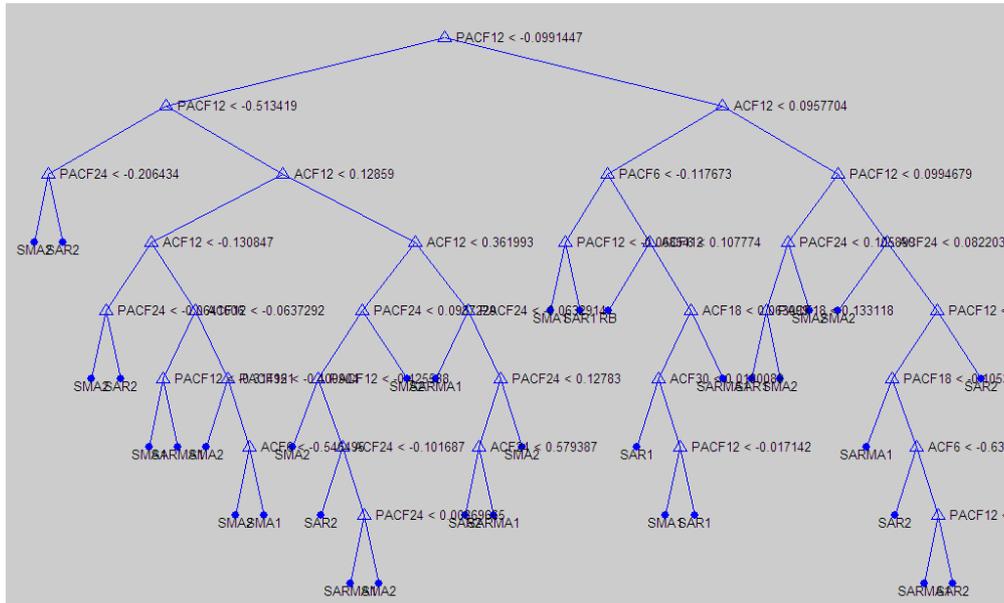


Figura 24 - Árvore de decisão para modelos sazonais de período 6

Para este conjunto de treinamento e teste, o erro no conjunto de teste é igual a 30.44% e o número ótimo de nós é de 67, sendo que 34 são nós terminais. O erro médio deste modelo é igual ao dos modelos não sazonais.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Árvore Modelo Sazonal Período 6

Performance Percentual no Treinamento

	% de Acerto	% de Erro
SAR1	89.5	10.5
SAR2	92	8
SMA1	87.1667	12.8333
SMA2	89	11
SARMA1	78.5	21.5
RB	95.1667	4.8333
Média	88.5556	11.4444

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	77.0833	22.9167
SAR2	74.3056	25.6944
SMA1	68.0556	31.9444
SMA2	68.75	31.25
SARMA1	37.5	62.5
RB	91.6667	8.3333
Média	69.5602	30.4398

➤ **Em relação à sazonalidade de período 12**

- Em relação ao conjunto de treinamento e teste inicial.

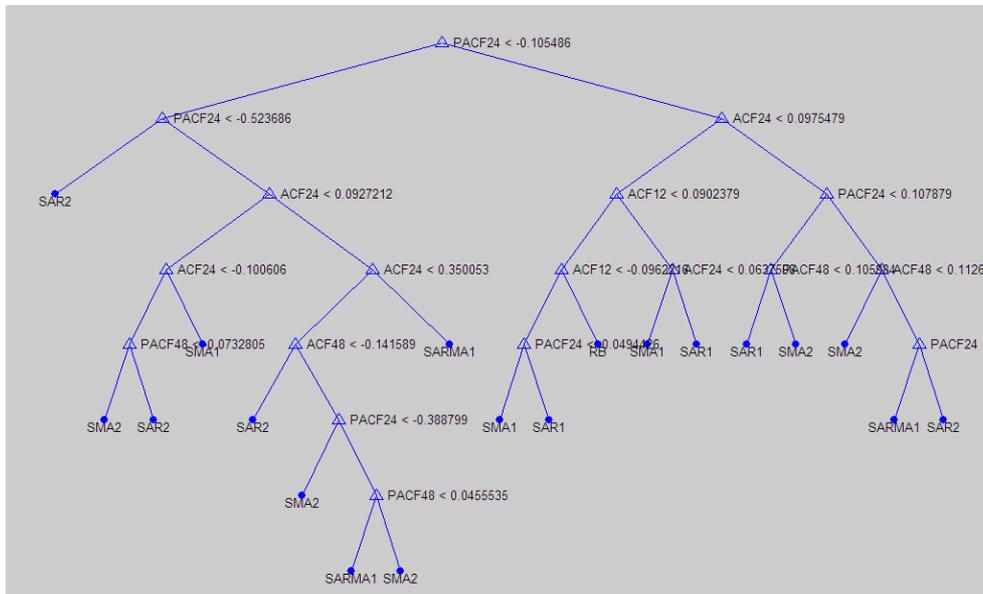


Figura 25 - Árvore de decisão para modelos sazonais de período 12

Para este conjunto de treinamento e teste, o erro no conjunto de teste é igual a 33.56% e o número ótimo de nós é de 37, sendo que 19 são nós terminais.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Árvore Modelo Sazonal Período 12

Performance Percentual no Treinamento

	% de Acerto	% de Erro
SAR1	87.8333	12.1667
SAR2	89.1667	10.8333
SMA1	90.1667	9.8333
SMA2	90.6667	9.3333
SARMA1	78.3333	21.6667
RB	94.8333	5.1667
Média	88.5	11.5

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	75.6944	24.3056
SAR2	56.9444	43.0556
SMA1	82.6389	17.3611
SMA2	60.4167	39.5833
SARMA1	31.25	68.75
RB	91.6667	8.3333
Média	66.4352	33.5648

6.5.2**Resultados obtidos ao utilizar Redes Neurais****6.5.2.1****Análise dos resultados para os modelos Box & Jenkins não-sazonais**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Rede Neural Modelo Não-sazonal

Performance Percentual no Treinamento

	% de Acerto	% de Erro
AR1	73	27
AR2	88.6667	11.3333
MA1	80.3333	19.6667
MA2	88.3333	11.6667
ARMA1	64.8333	35.1667
RB	95	5
Media	81.6944	18.3056

Performance Percentual no Teste

	% de Acerto	% de Erro
AR1	66.6667	33.3333
AR2	72.2222	27.7778
MA1	61.1111	38.8889
MA2	62.5	37.5
ARMA1	37.5	62.5
RB	81.9444	18.0556
Média	63.6574	36.3426

6.5.2.2**Análise dos resultados para os modelos Box & Jenkins sazonais**➤ **Em relação à sazonalidade de período 3**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Rede Neural Modelo Sazonal Período 3

Performance Percentual no Treinamento

	% de Acerto	% de Erro
SAR1	77.3333	22.6667
SAR2	88.3333	11.6667
SMA1	81	19
SMA2	84.3333	15.6667
SARMA1	70.1667	29.8333
RB	95.8333	4.1667
Média	82.8333	17.1667

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	62.5	37.5
SAR2	75	25
SMA1	60.4167	39.5833
SMA2	71.5278	28.4722
SARMA1	45.8333	54.1667
RB	90.9722	9.0278
Média	67.7083	32.2917

➤ **Em relação à sazonalidade de período 4**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Rede Neural Modelo Sazonal Período 4

Performance Percentual no Treinamento

	% de Acerto	% de Erro
SAR1	69.8333	30.1667
SAR2	84.8333	15.1667
SMA1	73.3333	26.6667
SMA2	82.6667	17.3333
SARMA1	56.5	43.5
RB	92.6667	7.3333
Média	76.6389	23.3611

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	59.0278	40.9722
SAR2	72.9167	27.0833
SMA1	54.8611	45.1389
SMA2	65.2778	34.7222
SARMA1	50	50
RB	86.1111	13.8889
Média	64.6991	35.3009

➤ **Em relação à sazonalidade de período 6**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Rede Neural Modelo Sazonal Período 6

Performance Percentual no Treinamento

	% de Acerto	% de Erro
SAR1	71.3333	28.6667
SAR2	83	17
SMA1	69.6667	30.3333
SMA2	85.1667	14.8333
SARMA1	61.6667	38.3333
RB	91.1667	8.8333
Média	77	23

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	59.7222	40.2778
SAR2	71.5278	28.4722
SMA1	51.3889	48.6111
SMA2	74.3056	25.6944
SARMA1	40.2778	59.7222
RB	90.2778	9.7222
Média	64.5833	35.4167

➤ **Em relação à sazonalidade de período 12**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros nos conjuntos de treinamento e teste.

Rede Neural Modelo Sazonal Período 12

Performance Percentual no Treinamento

	% de Acerto	% de Erro
SAR1	87.3563	12.6437
SAR2	90.1861	9.8139
SMA1	87.1094	12.8906
SMA2	87.1575	12.8425
SARMA1	67.1004	32.8996
RB	99.5	0.5
Média	86.4016	13.5984

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	72.8682	27.1318
SAR2	71.5278	28.4722
SMA1	76.5625	23.4375
SMA2	68.0851	31.9149
SARMA1	37.2093	62.7907
RB	97.2222	2.7778
Média	70.5792	29.4208

6.5.3**Resultados obtidos ao utilizar KNN – distância Euclidiana****6.5.3.1****Análise dos resultados para os modelos Box & Jenkins não-sazonais**

- o Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 8$.

KNN Modelo Não-sazonal - Distância Euclidiana

Performance Percentual no Teste

	% de Acerto	% de Erro
AR1	82.6389	17.3611
AR2	77.0833	22.9167
MA1	76.3889	23.6111
MA2	68.0556	31.9444
ARMA1	35.4167	64.5833
RB	95.1389	4.8611
Média	72.4537	27.5463

6.5.3.2**Análise dos resultados para os modelos Box & Jenkins sazonais**➤ **Em relação à sazonalidade de período 3**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 10$.

KNN Modelo Sazonal Período 3 - Distância Euclidiana

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	77.0833	22.9167
SAR2	75.6944	24.3056
SMA1	76.3889	23.6111
SMA2	72.9167	27.0833
SARMA1	42.3611	57.6389
RB	97.9167	2.0833
Média	73.7269	26.2731

➤ **Em relação à sazonalidade de período 4**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 7$.

KNN Modelo Sazonal Período 4 - Distância Euclidiana

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	76.3889	23.6111
SAR2	75.6944	24.3056
SMA1	74.3056	25.6944
SMA2	74.3056	25.6944
SARMA1	43.75	56.25
RB	90.2778	9.7222
Média	72.4537	27.5463

➤ **Em relação à sazonalidade de período 6**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 9$.

KNN Modelo Sazonal Período 6 - Distância Euclidiana

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	78.4722	21.5278
SAR2	73.6111	26.3889
SMA1	75	25
SMA2	77.0833	22.9167
SARMA1	34.7222	65.2778
RB	96.5278	3.4722
Média	72.5694	27.4306

➤ **Em relação à sazonalidade de período 12**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 8$.

KNN Modelo Sazonal Período 12 - Distância Euclidiana

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	79.8611	20.1389
SAR2	71.5278	28.4722
SMA1	75.6944	24.3056
SMA2	70.8333	29.1667
SARMA1	32.6389	67.3611
RB	98.6111	1.3889
Média	71.5278	28.4722

6.5.4

Resultados obtidos ao utilizar KNN – distância Manhattan

6.5.4.1

Análise dos resultados para os modelos Box & Jenkins não-sazonais

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 7$.

KNN Modelo Não-sazonal – distância de Manhattan

Performance Percentual no Teste

	% de Acerto	% de Erro
AR1	81.25	18.75
AR2	74.3056	25.6944
MA1	75	25
MA2	68.75	31.25
ARMA1	41.6667	58.3333
RB	92.3611	7.6389
Média	72.2222	27.7778

➤ **Em relação à sazonalidade de período 3**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 9$.

KNN Modelo Sazonal Período 3 – distância de Manhattan

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	77.0833	22.9167
SAR2	73.6111	26.3889
SMA1	74.3056	25.6944
SMA2	72.9167	27.0833
SARMA1	39.5833	60.4167
RB	97.9167	2.0833
Média	72.5694	27.4306

➤ **Em relação à sazonalidade de período 4**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 13$.

KNN Modelo Sazonal Período 4 – distância de Manhattan

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	75	25
SAR2	71.5278	28.4722
SMA1	77.0833	22.9167
SMA2	75.6944	24.3056
SARMA1	39.5833	60.4167
RB	93.75	6.25
Média	72.1065	27.8935

➤ **Em relação à sazonalidade de período 6**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 7$.

KNN Modelo Sazonal Período 6 – distância de Manhattan

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	77.0833	22.9167
SAR2	70.8333	29.1667
SMA1	72.9167	27.0833
SMA2	78.4722	21.5278
SARMA1	34.7222	65.2778
RB	97.2222	2.7778
Média	71.875	28.125

➤ **Em relação à sazonalidade de período 12**

- Em relação ao conjunto de treinamento e teste inicial.

Abaixo tem-se o percentual de acertos e erros no conjunto de teste onde $k = 9$.

KNN Modelo Sazonal Período 12 – distância de Manhattan

Performance Percentual no Teste

	% de Acerto	% de Erro
SAR1	77.0833	22.9167
SAR2	68.0556	31.9444
SMA1	79.8611	20.1389
SMA2	70.8333	29.1667
SARMA1	29.8611	70.1389
RB	97.2222	2.7778
Média	70.4861	29.5139

6.6 Conclusões

Com a finalidade de comparar as metodologias para a identificação automática dos Modelos Box & Jenkins, usou-se como critério de avaliação o percentual de acerto dos classificadores propostos: árvore de decisão, redes neurais, KNN e o obtido pelo *software Forecast Pro for Windows* (FPW) - Versão 3.50, em todos os conjuntos de teste propostos neste trabalho.

Na tabela a seguir, tem-se o percentual de acertos do conjunto de teste das séries não sazonais e sazonais utilizando o classificador **árvore de decisão**.

	Não-sazonal	Sazonal período 3	Sazonal período 4	Sazonal período 6	Sazonal período 12	Média
AR(1) ou SAR(1)	79,17	68,06	71,53	77,08	75,69	74,31
AR(2) ou SAR(2)	61,11	70,83	66,67	74,31	56,94	65,97
MA(1) ou SMA(1)	75,7	79,86	73,61	68,06	82,64	75,97
MA(2) ou SMA(2)	67,36	67,36	67,36	68,75	60,41	66,25
ARMA(1,1) ou SARMA(1,1)	42,36	38,19	36,11	37,5	31,25	37,08
Ruído Branco	91,67	95,83	93,75	91,67	91,67	92,92
Média	69,56	70,02	68,17	69,56	66,44	68,75

Tabela 2 - Resultados obtidos utilizando o classificador **árvore de decisão**

Na tabela 3 a seguir, tem-se o percentual de acertos do conjunto de teste das séries não sazonais e sazonais utilizando o classificador **redes neurais**.

	Não-sazonal	Sazonal período 3	Sazonal período 4	Sazonal período 6	Sazonal período 12	Média
AR(1) ou SAR(1)	66,67	62,5	59,03	59,72	72,87	64,16
AR(2) ou SAR(2)	72,22	75	72,92	71,53	71,53	72,64
MA(1) ou SMA(1)	61,11	60,42	54,86	51,39	76,56	60,87
MA(2) ou SMA(2)	62,5	71,53	65,28	74,31	68,09	68,34
ARMA(1,1) ou SARMA(1,1)	37,5	45,83	50	40,28	37,21	42,16
Ruído Branco	81,94	90,97	86,11	90,28	97,22	89,3
Média	63,66	67,71	64,7	64,58	70,58	66,25

Tabela 3 - Resultados obtidos utilizando o classificador **redes neurais**

Na tabela 4 a seguir, tem-se o percentual de acertos do conjunto de teste das séries não sazonais e sazonais utilizando o classificador **KNN – distância Euclidiana**.

	Não-sazonal	Sazonal período 3	Sazonal período 4	Sazonal período 6	Sazonal período 12	Média
AR(1) ou SAR(1)	82,64	77,08	76,39	78,47	79,86	78,89
AR(2) ou SAR(2)	77,08	75,7	75,7	73,61	71,53	74,72
MA(1) ou SMA(1)	76,39	76,39	74,31	75	75,7	75
MA(2) ou SMA(2)	68,06	72,92	74,31	77,08	70,83	72,64
ARMA(1,1) ou SARMA(1,1)	35,42	42,36	43,75	34,72	32,64	37,78
Ruído Branco	95,14	97,92	90,28	96,53	98,61	95,7
Média	72,45	73,73	72,45	72,57	71,53	72,46

Tabela 4 - Resultados obtidos utilizando o classificador **KNN – distância Euclidiana**

Na tabela 5 a seguir, tem-se o percentual de acertos do conjunto de teste das séries não sazonais e sazonais utilizando o classificador **KNN – distância de Manhattan**.

	Não-sazonal	Sazonal período 3	Sazonal período 4	Sazonal período 6	Sazonal período 12	Média
AR(1) ou SAR(1)	80,56	68,75	65,97	66,67	67,36	69,86
AR(2) ou SAR(2)	76,39	68,75	61,81	66,67	56,95	66,11
MA(1) ou SMA(1)	77,78	80,56	78,47	83,34	77,78	79,59
MA(2) ou SMA(2)	55,56	59,03	54,86	53,47	54,86	55,56
ARMA(1,1) ou SARMA(1,1)	36,12	39,58	38,89	31,95	27,08	34,72
Ruído Branco	90,28	97,23	91,67	90,97	98,61	93,75
Média	69,45	68,98	65,28	65,51	63,77	66,6

Tabela 5 - Resultados obtidos utilizando o **software FPW**

Os resultados obtidos por meio das três metodologias foram superados por aqueles encontrados pelo software FPW-XE.

Na tabela abaixo encontra-se a porcentagem média de acerto apresentada em cada metodologia.

Metodologia	Porcentagem média de acerto
Árvores de decisão	68,75
Redes Neurais	66,25
KNN – distância Euclidiana	72,46
KNN – distância de Manhattan	71,85
FPW - XE	66,60

Tabela 6 - Porcentagem média de acerto em cada metodologia

Das três metodologias apresentadas, o KNN foi o que apresentou melhor desempenho na identificação dos modelos Box & Jenkins, seguido da árvore de decisão e por último, as redes neurais.

A partir de agora, toda análise será realizada em função do classificador KNN – distância Euclidiana e do FPW. Nos dois classificadores, o percentual de acerto das estruturas ARMA(1,1) e SARMA(1,1) não foram significativos e este resultado deve ser analisado à parte, aplicando o teste de sobrefixação (capítulo 2) para detectar os parâmetros extras.

Dos erros gerados a partir das estruturas ARMA(1,1) e SARMA(1,1), verificou-se qual foi a classificação encontrada e a partir daí, quais poderiam representar um erro de sobrefixação que pode ser classificado como um erro bom. Nas tabelas abaixo tem-se o percentual encontrado das estruturas AR(1) e MA(1) para as estruturas ARMA(1,1) e o percentual encontrado das estruturas SAR(1) e SMA(1) para as estruturas SARMA(1,1) no KNN e FPW.

	Não-sazonal	Sazonal período 3	Sazonal período 4	Sazonal período 6	Sazonal período 12
AR(1) ou SAR(1)	23,61	17,36	20,83	21,53	26,38
MA(1) ou SMA(1)	20,83	22,22	21,53	20,83	24,31
Total	44,44	39,58	42,36	42,36	50,69

Tabela 7 - Percentual de AR(1), MA(1), SAR(1) e SMA(1) encontrados no ARMA(1,1) e SARMA(1,1) utilizando o KNN

	Não-sazonal	Sazonal período 3	Sazonal período 4	Sazonal período 6	Sazonal período 12
AR(1) ou SAR(1)	17,36	14,58	14,58	11,81	19,44
MA(1) ou SMA(1)	14,58	18,75	23,61	19,44	18,05
Total	31,94	33,33	38,19	31,25	37,49

Tabela 8 - Percentual de AR(1), MA(1), SAR(1) e SMA(1) encontrados no ARMA(1,1) e SARMA(1,1) utilizando o FPW

Da tabela 4, tem-se que o percentual de acerto da estrutura ARMA(1, 1) utilizando o KNN é de 35,42% e após a análise do número de AR(1) e MA(1) classificados (tabela 7), o percentual de acertos passa a ser de 79,86%. O mesmo processo é aplicado nos outros modelos para encontrar os demais percentuais de acerto, conforme percebido na tabela 9.

Na tabela 9 tem-se o percentual total de acerto das estruturas ARMA(1, 1) e SARMA(1, 1), após o teste de sobrefixação utilizando o classificador KNN.

	Não-sazonal	Sazonal período 3	Sazonal período 4	Sazonal período 6	Sazonal período 12
ARMA(1,1) ou SARMA(1,1)	35,42	42,36	43,75	34,72	32,64
AR(1) ou SAR(1)	23,61	17,36	20,83	21,53	26,38
MA(1) ou SMA(1)	20,83	22,22	21,53	20,83	24,31
Total	79,86	81,94	86,11	77,08	83,33

Tabela 9 - Total percentual de acertos encontrados no ARMA(1,1) e SARMA(1,1) utilizando o KNN

Da tabela 5, tem-se que o percentual de acerto da estrutura ARMA(1, 1) utilizando o FPW é de 36,12% e após a análise do número de AR(1) e MA(1) classificados (tabela 8), o percentual de acertos passa a ser de 68,06%. O mesmo processo é aplicado nos outros modelos para encontrar os demais percentuais de acerto, conforme percebido na tabela 10.

Na tabela 10 tem-se o percentual total de acerto das estruturas ARMA(1,1) e SARMA(1,1), após o teste de sobrefixação utilizando o FPW.

	Não-sazonal	Sazonal período 3	Sazonal período 4	Sazonal período 6	Sazonal período 12
ARMA(1,1) ou SARMA(1,1)	36,12	39,58	38,89	31,95	27,08
AR(1) ou SAR(1)	17,36	14,58	14,58	11,81	19,44
MA(1) ou SMA(1)	14,58	18,75	23,61	19,44	18,05
Total	68,06	72,91	77,08	63,2	64,57

Tabela 10 - Total percentual de acertos encontrados no ARMA(1,1) e SARMA(1,1) utilizando o FPW

6.7 Validação do classificador KNN

Para verificar a validação do classificador KNN foram geradas séries de alguns modelos SARMA(p,q)x(P,Q) e estas séries foram utilizadas no classificador KNN e FPW para a identificação dos parâmetros p, q, P e Q. Na tabela abaixo tem-se os modelos de séries geradas e sua respectiva identificação no KNN e FPW.

Modelo		Identificação KNN	Erro na identificação	Identificação FPW	Erro na identificação
1	SARMA(1,0)x(1,0)	SARMA(2,0)x(1,0)	X (**)	SARMA(2,0)x(1,0)	X (**)
2	SARMA(0,1)x(0,1)	SARMA(0,1)x(0,1)		SARMA(0,1)x(0,1)	
3	SARMA(1,1)x(1,1)	SARMA(1,1)x(1,1)		SARMA(1,1)x(1,0)	X (*)
4	SARMA(0,1)x(1,1)	SARMA(0,1)x(1,1)		SARMA(0,1)x(1,0)	X (*)
5	SARMA(1,0)x(1,1)	SARMA(1,0)x(1,1)		SARMA(1,0)x(1,1)	
6	SARMA(0,1)x(1,0)	SARMA(0,1)x(1,0)		SARMA(0,1)x(1,0)	
7	SARMA(1,0)x(0,1)	SARMA(1,0)x(0,1)		SARMA(1,0)x(0,1)	
8	SARMA(2,0)x(1,0)	SARMA(2,0)x(1,0)		SARMA(2,0)x(1,0)	
9	SARMA(2,0)x(2,0)	SARMA(2,0)x(0,1)	X (**)	SARMA(2,0)x(2,0)	
10	SARMA(2,0)x(1,1)	SARMA(1,0)x(1,1)	X (*)	SARMA(1,0)x(1,1)	X (*)
11	SARMA(0,2)x(0,1)	SARMA(0,2)x(0,2)	X (**)	SARMA(0,2)x(0,1)	
12	SARMA(0,2)x(0,2)	SARMA(0,2)x(0,1)	X (*)	SARMA(0,2)x(0,1)	X (*)

Tabela 11 - Classificação de alguns modelos Box & Jenkins

6.8

Validação do KNN utilizando a série *AIRLINE*

Agora para testar o sistema proposto, será utilizada uma série real apresentada em Box & Jenkins (1994:547). Esta série é composta do número mensal de passageiros em vôos internacionais (em milhares) de janeiro de 1949 a dezembro de 1960 e é conhecida como série *AIRLINE*. Ao todo tem-se 144 observações.

Segundo Box & Jenkins (1994:333), o modelo representativo desta série é SARIMA(0,1,1)x(0,1,1)₁₂. Como o sistema proposto neste trabalho não identifica a diferenciação de uma série, esta diferenciação não-sazonal e sazonal foi realizada a parte.

Abaixo tem-se a representação gráfica desta série. Pode-se observar que a série não é estacionária, isto é, a média e a variância não são constantes.

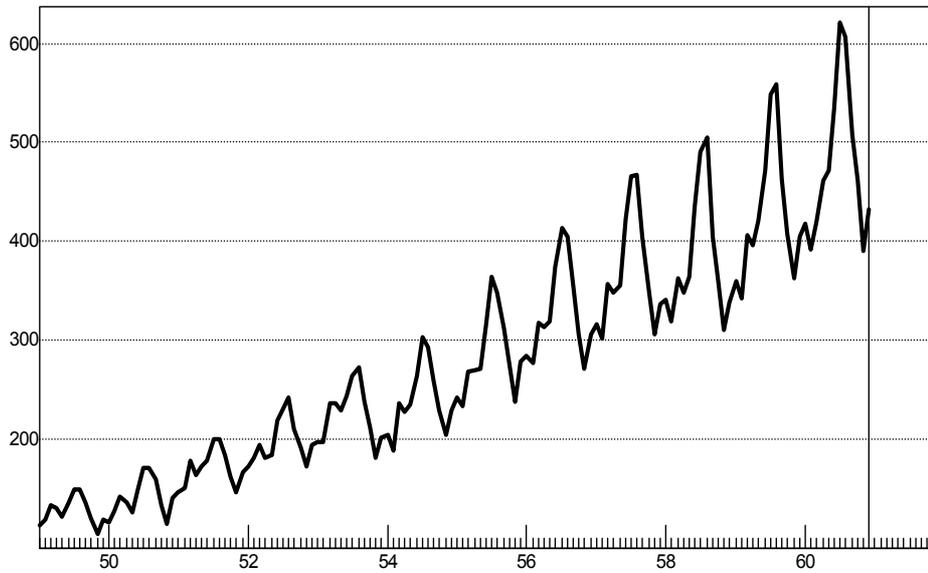


Figura 26 - Gráfico da série AIRLINE antes da diferenciação

Após a diferenciação, a série tornou-se estacionária como pode ser visto na figura abaixo.

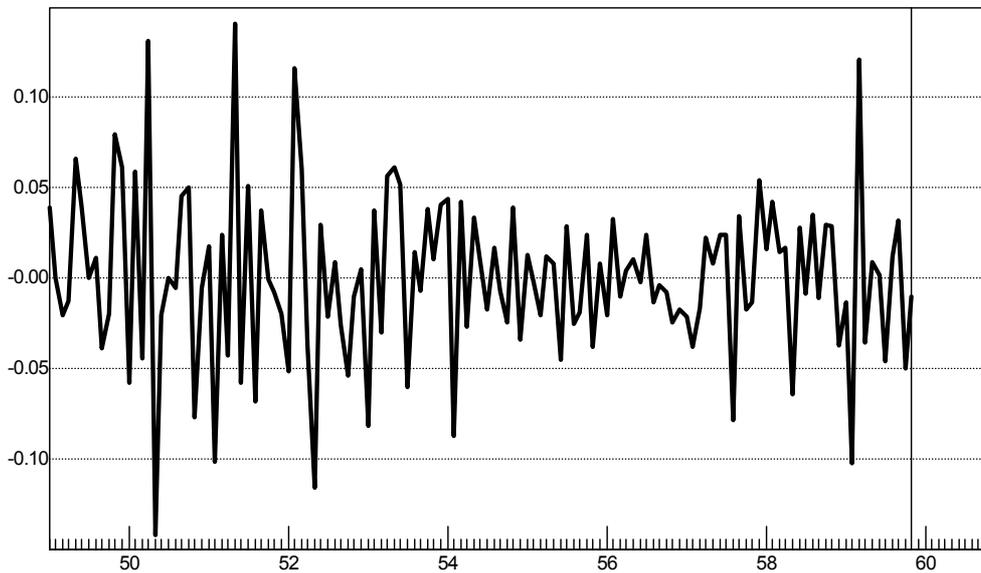


Figura 27 - Gráfico da série AIRLINE após a diferenciação

Feito a diferenciação, foi aplicado o classificador KNN a esta série e o resultado encontrado foi o do modelo SARMA(0,1)x(0,1)₁₂, como era o esperado.

Aplicando esta série no FPW, a identificação foi também correta, isto é, o modelo gerado foi SARMA(0,1)x(0,1)₁₂.